This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Increasing resolution and light eciency in uorescence microscopy

Wicker, Kai

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. May. 2023

Increasing resolution and light efficiency in fluorescence microscopy

Kai Wicker

King's College London
Randall Division of Cell and Molecular Biophysics

October 2010

A thesis submitted for the degree of Doctor of Philosophy at
King's College London

ABSTRACT

Structured illumination microscopy (SIM) can be used to enhance the lateral resolution and the sectioning capability in microscopic imaging. As a wide-field technique it may have advantages over scanning approaches such as stimulated emission depletion (STED) microscopy regarding acquisition time and bleaching. Similar to the generation of moiré patterns, the structured illumination transposes high spatial frequency information in the sample into low spatial frequency information, which can then be recorded by the microscope and computationally recovered in order to generate high resolution images. However, numerous experimental factors can lead to the generation of artefacts in the final image. Therefore, sophisticated algorithms are needed for the reconstruction of high quality images from raw data. This thesis discusses the experimental challenges of structured illumination microscopy and newly developed methods of addressing these in the reconstruction process. The final images show good quality and a resolution improvement of about a factor of two.

Confocal microscopy is another technique used for achieving optical sectioning of biological specimen. By using a pinhole in the detection pathway out-of-focus light is blocked, leading to the desired sectioning effect. Closing the pinhole further also allows an enhancement of the lateral resolution; however this comes at the cost of strongly reduced light efficiency, as less light passes through the pinhole. Adding an interferometer with relative image inversion optics between its two arms to the descanned output of a confocal microscope makes it possible to surpass the lateral resolution limit (closed pinhole) of confocal microscopes for large pinholes, significantly increasing the light efficiency of such a microscope. This thesis presents the theoretical description of this method as well as experimental results confirming it.

NOTE

In this version of the thesis Figure 6.1 on page 81 was changed in order to correct an error in the original thesis.

Contents

List of Figures

ACKNOWLEDGEMENTS

I would like to thank Rainer Heintzmann not only for the opportunity to do my PhD in his group, but also for giving me independence and freedom to pursue my own ideas, while at the same time being a great mentor.

Liisa Hirvonen, whose PhD work in Rainer's group overlapped with mine both temporally and thematically, I would like to thank for her experimental work on and productive discussions about structured illumination microscopy and also for proof-reading my thesis.

Ondrej Mandula played a great part in the development of the reconstruction algorithms and I would like to thank him for his input.

Thanks to Mats Gustafsson for many interesting discussions about structured illumination, which made some parts of the reconstruction process a lot clearer and for fantastic times spent visiting his group at *Janelia Farm*.

Enno Oldewurtel and Marie Walde were very helpful testing our *Zeiss* SIM prototype and trying to acquire nonlinear SIM data. I would especially like to thank Marie for her data I am using in this thesis.

Thanks to all my proof readers, but particularly Klaus Suhling (my second supervisor), Kristyna Pluhackova, Liisa and Rainer, who put the most work into it and had many useful comments.

Erith Yacht Club provided me with office space and some of the most amazing views of the river Thames making the last weeks of writing my thesis much more enjoyable, for which I am very grateful.

My parents as well as my uncles, Raimund and Micha, made it possible for me to study physics in the first place and I would like to thank them for their support,

especially my parents for their encouragement and support in everything I ever did.

Lastly, I want to thank my girlfriend, Katze, for being patient with me through the stressful times of this thesis, for taking my mind off my work when I needed it and for continuously showing me that there is more to life. Thank you for – *you*!

ABOUT THIS THESIS

This thesis contains work I did in the *NanoImaging Group* of Rainer Heintzmann at the *Randall Division of Cell and Molecular Biophysics, King's College London* between 2006 and 2010. It focuses on developments of two different techniques for increasing resolution and efficiency of fluorescence microscopy. It consists of three parts and an appendix.

Part I – Introduction and background gives a brief overview of several techniques available to the field of fluorescence microscopy, their benefits an shortcomings, and thus tries to provide a motivation for the research detailed in parts II and III. Furthermore, this part gives a very brief introduction to some optical concepts, such as point spread functions and optical transfer functions. Like the rest of the thesis, this introduction is aimed at readers trained in the field of optics. Its purpose is a quick recapitulation of a few concepts that will be needed parts II and III of this thesis as well as laying out a common notation. Readers familiar with microscopy techniques as well as optical concepts may find it sufficient to skip this part and begin directly with part II.

Part II – Structured illumination microscopy discusses the technique of structured illumination microscopy (SIM) in the context of resolution enhancement. Unlike other, more direct methods this technique requires intense computational image reconstruction, the principles of which are explained in this part. Although complicated, image reconstruction is somewhat straightforward under ideal experimental conditions. If, however, experimental conditions deviate from these ideal conditions, the reconstruction of high resolution SIM images requires sophisticated algorithms. The development of such algorithms (jointly with Rainer Heintzmann and Ondrej Mandula) constituted a major part of my PhD work. Part II explains the details of these algorithms and

shows high resolution SIM images reconstructed using these algorithms.

Part III – Resolution and efficiency enhancement using image inversion interferometers describes a novel method of increasing the resolution and light efficiency of confocal microscopes. It details the theory behind the technique and shows point spread function measurements as well as first biological images acquired with an interferometer developed during the course of my work.

The *Appendix* contains mathematical derivations and more in-depth analyses of ideas and equations presented within the three major parts of the thesis. Although some of these derivations constitute a major part of the actual work carried out, I chose to move them from the main part of the thesis to the appendix for the sake of better legibility.

Part I

INTRODUCTION AND BACKGROUND

1. INTRODUCTION

Fluorescence microscopy has become an indispensable tool of modern biology – combined with fluorescent tagging it allows the observation of structures and function in living organisms with high specificity as well as contrast. This technique has had a huge impact on microscopy and biology, a fact that will have influenced the Nobel committee when they awarded Shimomura, Chalfie and Tsien the 2008 Nobel Prize in Chemistry "for the discovery and development of the green fluorescent protein, GFP".

1.1 Optical sectioning

One shortcoming of wide-field imaging is the presence of out-of-focus blur in the recorded images, a direct consequence of the "missing cone" in the wide-field optical transfer function (OTF, see 2.2). When imaging thick samples, this out-of-focus blur can lead to a drowning of the in-focus signal in background light, rendering the structures of interest invisible.

Many techniques have been developed to fill this missing cone and thus block out-of-focus light and achieve optical sectioning. The best-known and most-used technique is confocal microscopy, where scanned illumination in combination with a detection pinhole leads to the desired result [2]. Spinning disc systems [36, 37, 48] and their digital counterpart – the programmable array microscope (PAM) [24] – employ the confocal principle in a highly parallelised way to enhance the acquisition rate of the microscopes. Two- and multi-photon scanning microscopy exploits the nonlinear relationship between illumination intensity and absorption probability to achieve sectioning [7]: although all sample slices are illuminated by the same total amount of light, more light is absorbed in the focal plane, as here the local intensity of the scanned illumination beam is high-

est. Structured illumination microscopy (SIM) employs sample illumination with a structure, such as periodic grating [45] or point patterns [3]. The acquisition of three or more images with different illumination grating positions allows the computational reconstruction of one optically sectioned image (see 3.1). Selective plane illumination microscopy (SPIM) generates relatively thin sheets of light perpendicular to the optical axis of the system, and thus ensures that out-of-focus areas are not illuminated and therefore cannot generate out-of-focus blur [35, 61]. The same principle is used by oblique plane microscopy (OPM), the main difference being that in the latter the same objective is used for illumination as well as detection: a light sheet is generated at an oblique angle by illuminating an off-centre line in the back focal plane (BFP). For the detection a conjugate region of the BFP is used, giving the impression of looking perpendicularly onto the illumination sheet [9]. Yet another interesting approach is wide-field two-photon excitation imaging using temporal focusing (TF) [46]. In this method spatiotemporal pulse shaping is used to ensure that the pulsed excitation field is compressed as it propagates through the sample, resulting in maximum axial compression in the focal slice, where it will thus reach its peak intensity. Because of the two-photon excitation's quadratic dependence on the illumination intensity this leads to a wide-field sectioning effect.

1.2 Resolution enhancement

Another shortcoming of optical microscopy is the relatively low resolution, which in conventional wide-field systems is diffraction-limited to about 200 nm laterally and 500 nm axially, described by the Abbe-limit of resolution [1]. Other methods such as electron microscopy (EM) [52, 53] or atomic force microscopy (AFM) [6] can acquire images of much higher resolution (EM has been shown to achieve a resolution of 0.5Å [11]), but these methods cannot be used with living cells. Apart from the ability to image living cells, fluorescence microscopy has the further advantage that it can generate images which are essentially free of background and that fluorescent markers can specifically target molecules such as certain proteins of interest, allowing imaging

with high specificity.

Optical techniques employing inhomogeneous rather than full-field illumination can improve the resolution of optical microscopes, e.g. in confocal microscopy the total PSF is the product of the illumination PSF and the detection PSF. In Fourier space this corresponds to a convolution of the respective OTFs, yielding a total maximum support frequency (which defines the famous Abbe resolution limit [1]) that is the sum of the individual maximum support frequencies. Besides optical sectioning, SIM can also be used for resolution improvement [15, 22] (see 3.2). Similar to the moiré effect, the illumination with a high frequency grating pattern down-modulates high frequencies present in the sample to lower frequencies, which can then be captured by the microscope. Computational separation and up-modulation of this information to its true position in Fourier space extends the support of the imaging OTF and hence enhances the resolution. However, as the illumination is itself diffraction-limited, the maximum possible resolution enhancement is again about two-fold.

The axial resolution of confocal microscopes can further be enhanced by using two opposing objectives to coherently increase the solid angle of illumination (and detection) and therefore also the available spatial frequencies. The 4Pi microscope employs this technique. In its type A configuration the support of the illumination OTF is increased. In the type C configuration, also light collected through the two objectives is coherently recombined, increasing the support of the detection OTF in addition to that of the illumination OTF. This yields an axial resolution of about 100 nm. As a wide-field equivalent to the 4Pi, the I^5M achieves similar axial resolution [18].

1.3 Circumventing the Abbe-limit

In all of the above techniques, both illumination and detection remain diffraction-limited and the Abbe-limit is not broken [23]. This limit can however be circumvented and surpassed by employing nonlinearities in the illumination process: while the illumination pattern itself is diffraction limited, a nonlinear fluorescence response can result in fluorescence emission corresponding to an effective illumination containing spatial

frequencies outside the support of the illumination OTF. In SIM such nonlinearities can be achieved through the saturation of the fluorescence process [16] or the use of photo-switchable fluorophores [31, 32]. The resolution that can be achieved this way is in principle only limited by the signal-to-noise ratio (SNR) of the acquired images.

In stimulated emission depletion microscopy (STED) [29] the sample is excited using a focused laser beam, much like in confocal microscopy. A second, ring- or doughnut-shaped beam with a central intensity zero forces fluorophores inside the excitation volume back into the ground-state via the process of stimulated emission. Only at or very near the position of the depletion beam's intensity zero will the fluorophores remain in the excited state, significantly reducing the effective excitation volume. Any fluorescence light collected after the depletion process can therefore be attributed to this reduced excitation volume, which because of the nonlinearity of the depletion can in principle be made arbitrarily small. Besides stimulated emission, the same effect can also be achieved through the use of photo-switchable fluorophores [34]. To date a lateral resolution of 5.8 nm has been achieved for imaging crystal colour centres [51], and 15 nm in biological samples [8], a twelve-fold improvement over confocal microscopy.

Another approach is based on the fact that while a single fluorophore can only be imaged as a diffraction limited spot, its position can be localised with a much higher precision. Many fluorophores can then be localised consecutively to form an image composed of many individual points. The main prerequisite for this technique is the ability to assign the measured photons to individual fluorophores. This can be done through analysing the blinking statistics of quantum dots as in pointillism [40] or by repeatedly sparsely activating photo-switchable dyes, so that their images do not overlap, as in stochastic optical reconstruction microscopy (STORM) [54], photo-activated localization microscopy (PALM) [5] or fluorescence photoactivation localization microscopy (FPALM) [30].

1.4 Motivation of this thesis

SIM allows the imaging of biological samples with high resolution. While it is possible to build very precise SIM systems allowing a relatively straightforward image reconstruction, many simpler systems cannot guarantee the same stability and are therefore prone to imaging artefacts in the reconstructed images. Such systems can nevertheless produce high quality images if the reconstruction algorithms used can estimate and correct experimental fluctuations. The development of such algorithms – as detailed in part II – can therefore help to make this technique available to a wide range of users.

While there have been many developments in fluorescence microscopy, particularly in the field of high resolution imaging, the confocal microscope remains the workhorse of the biological imaging community. However, its main advantages (optical sectioning and resolution enhancement) come at the price of reduced signal-to-noise levels. Techniques increasing the efficiency and resolution of confocal microscopes are therefore of great interest. One such technique – image inversion interferometry – is presented in part III.

2. SOME BASIC OPTICAL CONCEPTS

2.1 Resolution criteria

When imaged with a microscope (or any other optical system) a point in the sample will not be resolved as a point in the image. Instead it will create a spot of finite size, called the *point spread function* (PSF) – its nature will be discussed in more detail in 2.2. Due to the size of this PSF, points that are too close to each other in the sample may no longer be distinguishable in the image, the distance between them is unresolvable. Therefore the resolution of an optical system is closely related to size and shape of its PSF.

Given the PSF of an optical system, there are several criteria that can be used to define its resolution capability. According to the *Sparrow criterion* [63] two points can be considered to be resolved, as long as their image (the sum of their two PSFs) exhibits a central dip. The *Rayleigh criterion* [49] defines two points as resolvable, if their distance is greater or equal to the distance between the PSF's maximum and first minimum.

While these criteria give some measure for the resolution capability of an optical system, they are not very useful, as their definitions are relatively arbitrary. But worse, these criteria do not define clear or absolute limitations for an optical system: even after an image has been recorded, the size and shape of a PSF can be altered retrospectively by applying digital filters. By attenuating low spatial frequencies and thus biasing the image towards high frequencies, the resolution according to the above criteria could be improved, while clearly neither additional sample information is actually gained through this process, nor is the physical quality of the system improved in any way.

A better way to define resolution is by means of the *optical transfer function* (OTF).

Through a Fourier transform (FT) an image can be decomposed into its various spatial frequency components. Fine detail requires high spatial frequencies, whereas a constant feature in the image (i.e. background) would be represented by the zero spatial frequency. The OTF is a measure for how well the different frequency components are transmitted through the optical system. It is the Fourier transform of the PSF and will also be discussed in more detail in 2.2. The resolution of an optical system is fundamentally limited by the *support* of the OTF, which defines a maximum spatial frequency range that can be transmitted. While attenuating lower frequencies can make the higher frequencies more dominant and hence improve detail in an image, it cannot increase the support of the OTF. Frequencies outside the support of the OTF are lost in the imaging process and cannot be recovered by conventional means.

The support of an OTF is closely related to the Abbe resolution limit [1]. This criterion states that sample structures can be resolved if the first order of light diffracted by them are captured by the optical system. A periodic structure such as e.g. a diffraction grating will diffract light at an angle β. If this angle is smaller than the acceptance angle of the optical system $\alpha = \arcsin(NA/n)$, (with NA being the numerical aperture and n the refractive index of the system) the light can be captured by the system and the structure can be resolved. This results in the Abbe resolution limit of $\Delta d_{min} = \lambda/2NA$.

2.2 Point spread function and optical transfer function

Any light distribution in the detector space can be described as a superposition of plane waves arriving under different angles. The higher the angle, the narrower the waves' modulation in the detector's lateral plane will be, leading to finer detail in a light/dark distribution. If the light is of a single wavelength only, the k-vectors of all plane waves will lie on a spherical shell centred around the origin in Fourier space, known as the Ewald sphere [12] (Fig. 2.1). The radius of the shell is defined by the wavelength of the light, the angle of the k-vector corresponds to the propagation angle of the plane wave.

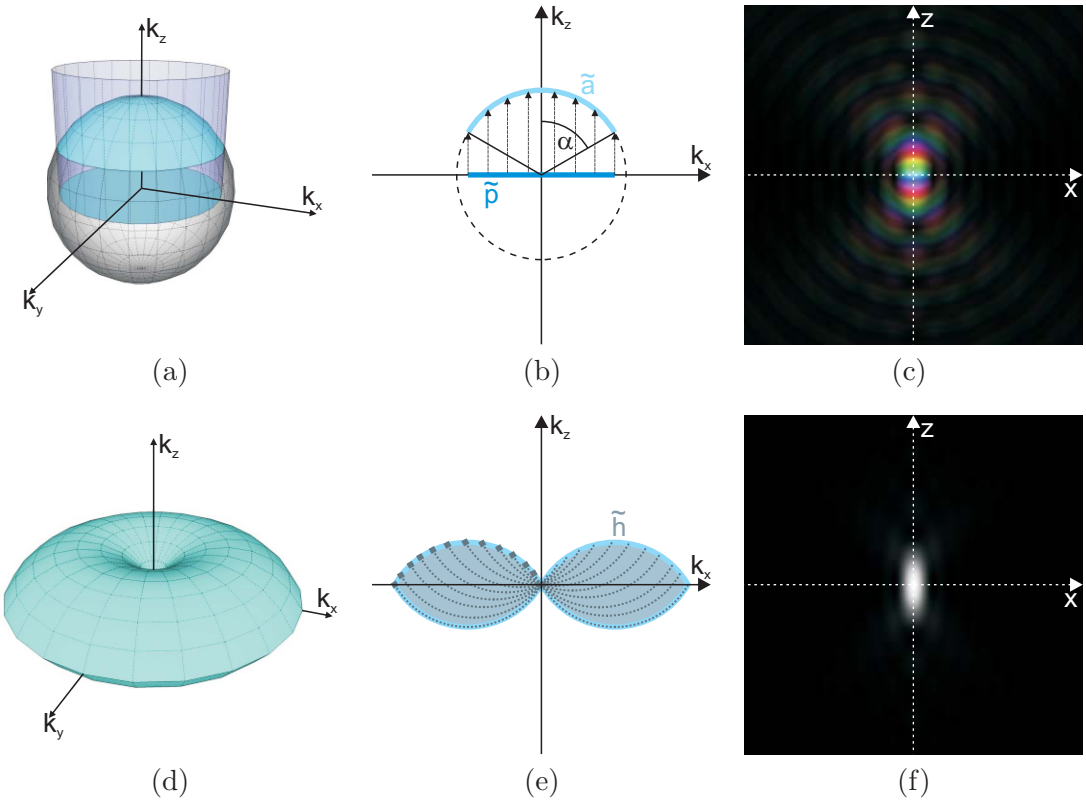


Fig. 2.1: (a) All k-vectors of waves of a single wavelength λ lie on a spherical shell. Projecting the optical system's pupil function $\tilde{p}(\vec{k}_{xy})$ (blue circle in the k_x, k_y-plane) onto the Ewald sphere, yields the k-vectors available for imaging. (b) A two-dimensional representation of the pupil function's projection onto the Ewald sphere, where α represents the acceptance angle of the microscope objective. This projection is the McCutchen generalised aperture. (c) An inverse Fourier transform of the generalised aperture yields the field distribution for a single point emitter in the sample space's origin, which is equivalent to the systems ATF. Brightness and colour represent amplitude and phase. (f) The intensity distribution (PSF) is given by the absolute square of the field distribution. The corresponding OTF is an autocorrelation of the ATF and is shown in (d) and (e). It has a "missing cone" of spatial frequencies around the k_z axis.

The k-vectors available for the formation of an image are defined by what angles can be captured by the optical system; the larger the numerical aperture (NA) of the objective, the larger the angles available. More precisely, the system's capability to capture light under different angles is described by its pupil function $\tilde{p}(\vec{k}_{xy})$. This function is 1 if the wave is transmitted unchanged, but it can also account for an

attenuation ($|\tilde{p}| < 1$) and phase changes introduced by the system ($\arg(\tilde{p}) \neq 0$). The projection of this pupil function onto the Ewald sphere (or McCutchen's generalised aperture [42]) defines the k-vectors contained in the field distribution and is equivalent to the coherent or amplitude transfer function (ATF) of the system. The inverse Fourier transform of this ATF corresponds to the field distribution generated in detector space by a single point situated in the origin of the sample space, which is equivalent to the system's coherent or amplitude point spread function (APSF). To get the intensity distribution, we need to take the absolute square of the field distribution, which gives us the PSF of the system, $h(\vec{r}) = |a(\vec{r})|^2 = a(\vec{r})a^*(\vec{r})$. A Fourier transform of this PSF yields the OTF $\tilde{h}(\vec{k})$, which similar to the ATF indicates the transfer capability of frequencies in an intensity picture.

2.3 Image formation

Light emanating from a single point in sample space can be described by the electric field distribution it generates in the detector plane. This can be expressed through the APSF $a(\vec{r} - \vec{r}')|_{z=0}$, where $\vec{r} = (x, y, z = 0)$ denotes detector space (scaled as to contain any necessary magnification factor), and $\vec{r}' = (x', y', z')$ is the point's position in sample space. As a detector records intensity rather than amplitude, the recorded image of the point source would be $h(\vec{r} - \vec{r}') = |a(\vec{r} - \vec{r}')|^2$.

2.3.1 Incoherent imaging

In incoherent imaging, e.g. for fluorescence or under incoherent illumination, light emanating from different points in the sample does not interfere. In order to get the final intensity distribution corresponding to imaging the whole sample rather than just one point, we need to sum the individual PSF images of all points, i.e. integrate over the sample's point distribution or sample response $S(\vec{r}')$:

$$I(\vec{r}) = \int S(\vec{r}')h(\vec{r} - \vec{r}')d^3r' = (S \otimes h)(\vec{r}). \tag{2.1}$$

This can only be done for systems with a linear shift invariant PSF. It can readily be seen that the above integral is a convolution of the sample distribution with the PSF, which is denoted by the operator $\otimes$. Structured illumination microscopy, which will be discussed in great detail in part II of this thesis relies heavily on image analysis in Fourier space. We shall therefore already now take a look at the Fourier transforms of the equations describing the image formation process. The above equation's FT is

$$\tilde{I}(\vec{k}) = \tilde{S}(\vec{k})\tilde{h}(\vec{k}). \tag{2.2}$$

Equations 2.1 and 2.2 assume homogeneous illumination. If the illumination instead is inhomogeneous, with a spatial dependence $L(\vec{r}')$ in sample space, it has to be accounted for. Equation 2.1 thus becomes:

$$\begin{aligned} I(\vec{r}) &= \int L(\vec{r}')S(\vec{r}')h(\vec{r}-\vec{r}')d^3r' = (\{LS\}\otimes h)(\vec{r}), \\ \tilde{I}(\vec{k}) &= \{\tilde{L}\otimes\tilde{S}\}(\vec{k})\tilde{h}(\vec{k}). \end{aligned} \tag{2.3}$$

However, this equation assumes a linear response of emitted light to the illumination. This is not always the case. For high illumination intensity for example, the fluorescence process becomes saturated. In the case of nonlinear response Eq. 2.3 can be written as

$$\begin{aligned} I(\vec{r}) &= \int S(L(\vec{r}'))h(\vec{r}-\vec{r}')d^3r' = (S(L)\otimes h)(\vec{r}), \\ \tilde{I}(\vec{k}) &= \widetilde{S(L)}(\vec{k})\tilde{h}(\vec{k}). \end{aligned}$$

Here $S(L)$ indicates that the sample response depends nonlinearly on the illumination intensity.

2.3.2 Coherent imaging

In many cases the imaging process can be coherent, e.g. for transmission or reflection microscopy with coherent illumination. In these cases it will not be enough to sum all

the points' intensity PSF images to get the final images. Because light coming from different points within the sample can interfere, one has to consider field strengths rather than intensities and therefore work with the coherent PSF or APSF $a(\vec{r})$. In the case of reflection the light field reflected by the sample is $l(\vec{r})s(\vec{r})$, where $l(\vec{r})$ is the illuminating light field ($L(\vec{r}) = |l(\vec{r})|^2$) and $s(\vec{r})$ the reflectance corresponding to sample's reflectivity $R(\vec{r}) = |s(\vec{r})|^2$. Both $l(\vec{r})$ and $s(\vec{r})$ describe field strengths rather than intensity and may contain phase terms. The final field distribution is described via a convolution with the APSF a. The detector still measures intensity and we therefore have to take the absolute square of the field, which yields

$$I_{\text{coherent}}(\vec{r}) = |\{[ls] \otimes a\}(\vec{r})|^2 .$$

Although all the techniques discussed in this thesis are incoherent imaging techniques, the interferometric technique discussed in part III requires an understanding of coherent imaging, as it is coherence and interference effects which lead to the improved PSF of this method.

Part II

STRUCTURED ILLUMINATION MICROSCOPY

3. INTRODUCING STRUCTURED ILLUMINATION MICROSCOPY

Parts of this chapter including figures have been published in [69] **K. Wicker and R. Heintzmann, *Single-shot optical sectioning using polarization-coded structured illumination*, Journal of Optics, Special Issue on Ultrafast Biophotonics (2010), in press**.

Structured illumination microscopy is a well-established technique for acquiring optically sectioned images similar to those of confocal microscopy [45]. For this a coarse grating is incoherently imaged into the sample. Three images taken under different grating positions can then be computationally recombined to reconstruct one optically sectioned image. The imaging properties of this technique will be discussed in 3.1 [69].

The strongest potential of SIM however lies in resolution enhancement. Through illumination with a periodic light pattern, high spatial frequency information of the sample is down-modulated to lower frequencies and thus made available for capture by the microscope. Linear SIM can improve the resolution by a factor of about two this way. Exploiting nonlinearities in the sample response to the illumination light (e.g. through fluorescence saturation, or by using photo-switchable dyes), nonlinear SIM is fundamentally limited in resolution only by the strengths of the nonlinearities and the signal-to-noise levels of the acquired data. Depending on the quality of the available data, sophisticated algorithms may be required for the reconstruction of resolution-improved images. This chapter will discuss such algorithms developed in our group in detail.

3.1 Structured illumination microscopy for optical sectioning

In sectioning SIM we can describe the illumination as a sinusoidal pattern which is incoherently imaged into the sample. Not considering partial coherence effects, this leads to an intensity distribution of

$$L_{\mathrm{illu},n}(\vec{r}) = L/2\{\delta(z)(1+\cos(\vec{k}_{\mathrm{g}}\vec{r}+2n\pi/3))\} \otimes h_{\mathrm{illu}}(\vec{r}), \tag{3.1}$$

with $\vec{r}$ denoting the sample coordinates, $\delta(z)$ the Dirac delta, $\vec{k}_{\mathrm{g}}$ the grating's k-vector in sample coordinates (which is assumed to have lateral components only), $h_{\mathrm{illu}}(\vec{r})$ the illumination point spread function, $\otimes$ the convolution operator and L the mean intensity. The index n denotes the image number and accounts for the grating displacement in the different images. In the sample the intensity pattern exhibits good contrast in the focal slice, but reduced contrast for out-of-focus regions, owing to out-of-focus blurring of the incoherent illumination. Therefore the amount of light illuminating a single point in the sample varies strongly with the grating position (index n) for in-focus points, but becomes more and more constant for greater distances from the focus.

The final two-dimensional (2D) images recorded on the camera are

$$I_n(\vec{r}_{xy}) = \{(L_{\mathrm{illu},n}S) \otimes h_{\mathrm{det}}\}(\vec{r})|_{z=0}\,,\ n=\{1,2,3\}, \tag{3.2}$$

where $S(\vec{r})$ is the sample's fluorophore density (assuming a linear response) and $h_{\mathrm{det}}(\vec{r})$ is the detection PSF. The vector $\vec{r}_{xy}$ denotes the 2D image coordinates. From these images a sectioned slice can be computed by [45]

$$I(\vec{r}_{xy}) = \sqrt{(I_1(\vec{r}_{xy})-I_2(\vec{r}_{xy}))^2+(I_2(\vec{r}_{xy})-I_3(\vec{r}_{xy}))^2+(I_3(\vec{r}_{xy})-I_1(\vec{r}_{xy}))^2},$$

which is mathematically identical to a form corresponding to homodyne detection [45],

$$I(\vec{r}_{xy}) = \left|\sum_{n=1}^{3} I_n(\vec{r}_{xy})\exp\{\imath 2\pi n/3\}\right|, \tag{3.3}$$

where $\imath = \sqrt{-1}$ is the imaginary unit.

This reconstruction suppresses the unmodulated components in the series of three images $I_n(\vec{r}_{xy})$ and keeps the modulated ones. The degree of modulation in the images that can be attributed to the illumination is strongest for the in-focus slice, whereas the loss of contrast in the illumination pattern for out-of-focus regions reduces the degree of modulation. Any modulation stemming from the patterned illumination may be further reduced in the recorded images if the PSF images of neighboring sample points (which may both be modulated but with a different phase) overlap. Because of blurring this overlap is greater for out-of-focus areas than for the focal slice. Therefore both incoherent illumination and out-of-focus blurring in the detection add to the sectioning effect.

As shown in appendix A.1, the reconstructed image (Eq. 3.3) can then be written as

$$I(\vec{r}_{xy}) = \left|([mS] \otimes h'_{\mathrm{det}})(\vec{r})\right|\big|_{z=0}\,. \tag{3.4}$$

Here $h'_{\mathrm{det}}(\vec{r})$ is the effective detection PSF (Fig. 3.1c)

$$h'_{\mathrm{det}}(\vec{r}) = F^{-1}_{\vec{k}}[\tilde{h}_{\mathrm{det}}(\vec{k} + \vec{k}_{\mathrm{g}})](\vec{r}) = h_{\mathrm{det}}(\vec{r}) \exp\{\imath \vec{k}_{\mathrm{g}} \vec{r}\}, \tag{3.5}$$

which contains spatial frequency information filling the missing cone. F^{-1} denotes the inverse Fourier transform with respect to the directions indicated in the index; $m(\vec{r})$ is the effective sample modulation in the direction of the optical axis (Fig. 3.1c, d):

$$m(\vec{r}) = m(z) = F^{-1}_{\vec{k}}[\delta(k_x)\delta(k_y)\tilde{h}_{\mathrm{illu}}(k_x - k_{\mathrm{g},x}, k_y - k_{\mathrm{g},y}, k_z)](\vec{r}). \tag{3.6}$$

After taking the modulus in Eq. 3.4, the image formation can no longer be expressed as a linear convolution of the sample with a PSF, as the imaging properties become sample dependent, e.g. neighboring sample structures may cancel each other out in the final image – a result of the convolution with the complex valued PSF $h'_{\mathrm{det}}(\vec{r})$ (Eq. 3.5). Consequently, the image reconstruction process also has no true OTF. However, the

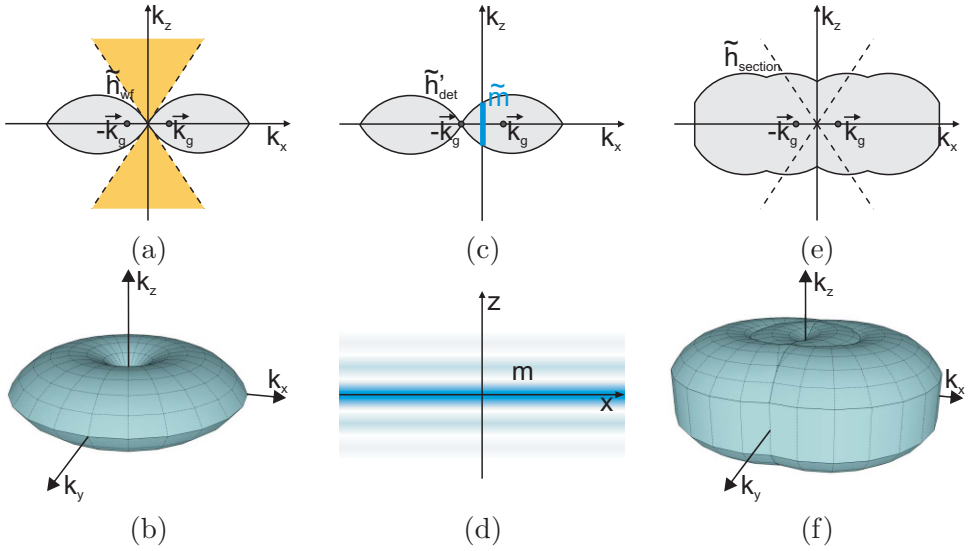


Fig. 3.1: The wide-field optical transfer function $\tilde{h}_{\mathrm{wf}}$ shown in (a) and (b) has a "missing cone" of spatial frequencies around the k_z direction (indicated in orange in (a)), leading to blurred out-of-focus light in the acquired wide-field images. The acquisition of three images under patterned illumination and subsequent computational demodulation (before taking the modulus) result in an effective detection OTF $\tilde{h}'_{\mathrm{det}}$ which is the wide-field OTF shifted by $-\vec{k}_{\mathrm{g}}$. As can be seen in (c) $\tilde{h}'_{\mathrm{det}}$ fills the missing cone and thus results in optically sectioned images. Furthermore, it also has an increased support in the direction of the grating's k-vector, leading to an increase in lateral resolution in that direction. Illumination with incoherent light further leads to an effective modulation of the sample with a function $m(z)$ (d), which is defined in Fourier space by the intersection of the $\vec{k}_{\mathrm{g}}$-shifted wide-field OTF (similar to $\tilde{h}'_{\mathrm{det}}$, but for illumination rather than detection) with the k_z axis (c). Taking the modulus symmetrises the support, whereas the modulation with $m(z)$ extends it in the k_z direction. However, the resulting final support shown in (e) and (f) only shows which frequencies are theoretically available. It is not a true OTF, as the final image cannot be written as a linear convolution of the sample with a PSF.

frequency information present in the final image will be from within the support of the OTF $\tilde{h}_{\mathrm{section}}(\vec{k}) = \tilde{m}(\vec{k}) \otimes (\tilde{h}'_{\mathrm{det}}(\vec{k}) + \tilde{h}'_{\mathrm{det}}(-\vec{k}))$ (Fig. 3.1e, f). This support fills the missing cone, and has an extended spatial frequency support in both the lateral and the axial direction.

On the other hand it is worth noting that *before* taking the modulus (Eq. 3.4) the reconstruction *is* in fact a linear convolution, albeit one with a complex valued PSF. The modulus then simply disregards the phases and looks only at the amplitude of

the complex image. This is contrary to the notion that this reconstruction is strongly nonlinear.

In analogy with HiLo microscopy [59] it can be advantageous for signal-to-noise reasons to low-pass filter images obtained through Eq. 3.3 and combine them with a high-pass filtered wide-field image $I_{\mathrm{wf}} = I_1 + I_2 + I_3$.

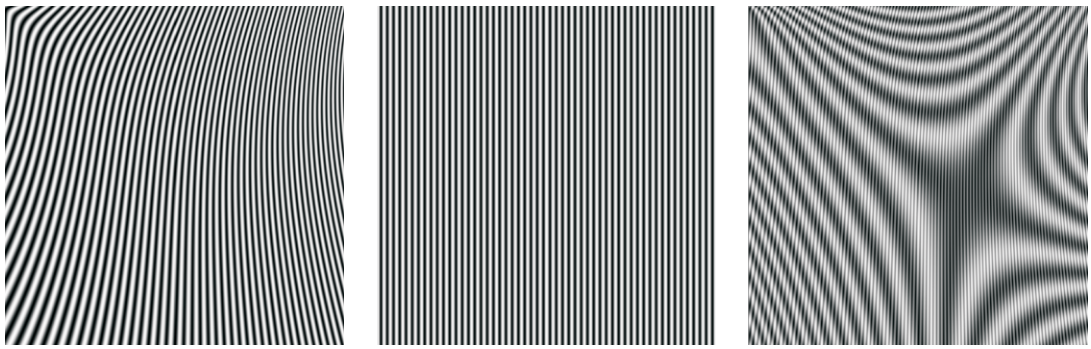
The reconstructed SIM images may exhibit artefacts stemming from aberrations in the imaging optics, intensity fluctuations, photo-bleaching, illumination with non-sinusoidal gratings, etc. However, many of these effects can be corrected for algorithmically [60].

3.2 Structured illumination microscopy for resolution improvement

Besides filling the missing cone, sectioning SIM increases the OTF support in the lateral as well as the axial direction. In addition to optical sectioning it therefore also yields an improved resolution. As the illumination pattern used in sectioning SIM is usually chosen to be quite coarse, the resolution improvement is usually quite minute. Using finer gratings opens the possibility for significant resolution improvement.

3.2.1 The moiré effect

When a sample is illuminated with patterned illumination, high spatial frequency sample information (from outside the microscope's OTF support) is down-modulated to lower frequencies, which can then be captured by the microscope. This down-modulation is known as the moiré effect which can be readily observed in everyday life, e.g. in the brightness changes seen when looking through two fences. Figure 3.2 illustrates the moiré effect: multiplying a high frequency structure (Fig. 3.2a) with a high frequency periodic grating (Fig. 3.2b) results in an intensity distribution containing both high and low frequencies (Fig. 3.2c). If only the low frequencies are captured by an optical system, knowledge of the grating pattern can be used to derive the original high frequency structure.



(a) (b) (c)

Fig. 3.2: The moiré effect. When a high-frequency pattern (a) is multiplied with another high frequency, regular pattern (b), the resulting pattern will exhibit both high and low frequencies (c). Whereas high frequencies may not be captured in an imaging process due to the limited NA of the system, the low frequencies may well be detectable. Combined with knowledge of the regular high-frequency pattern, these low frequencies can be used to deduce information contained in the first high-frequency pattern, which would otherwise be lost in the imaging process.

3.2.2 How SIM enhances resolution

Structured illumination microscopy uses this effect to recover high spatial frequency information that would otherwise be lost in the imaging process. Figure 3.3 shows a graphical illustration of the principle of high resolution SIM. The synthetic sample (Fig. 3.3a1) exhibits fine detail, which corresponds to high spatial frequencies in its Fourier space representation (Fig. 3.3a2). Note the light blue boxes in the bottom right corner of the real space images, which show a close-up of image detail. Imaging with an ordinary microscope leads to high frequencies being cut off (Fig. 3.3b2) because of the limited support of the system's OTF (shown as a red dotted circle). This leads to a loss of detail and a blurring of the image (Fig. 3.3b1). The illumination of the sample with a sinusoidal grating pattern (Fig. 3.3c1), which in Fourier space corresponds to three peaks (Fig. 3.3c2), can be thought of as a multiplication of sample and grating (Fig. 3.3d1). In Fourier space this multiplication corresponds to a convolution of the Fourier representations of sample and grating. This can be thought of as a sum of three copies of the Fourier representation of the sample, positioned at each of the three grating peaks (Fig. 3.3d2, e1). When imaged, once again only frequencies inside the

support of the OTF are captured (Fig. 3.3e1). However, this information now contains sample information from different (higher) frequency regions. The acquisition of at least three images under different grating positions provides enough information to be able to mathematically separate these three components (Fig. 3.3e1). Shifting them so that their true zero frequency is aligned with the zero frequency in Fourier space these components can be combined (Fig. 3.3e2), yielding an effective OTF which has been extended to frequencies beyond the support of the original OTF (Fig. 3.3f2). This leads to an enhancement of the resolution in this direction (Fig. 3.3f1). This process has to be repeated for different grating orientations in order to achieve near isotropic lateral resolution enhancement. Although Fig. 3.3 illustrates only the lateral resolution improvement of SIM, the technique also enhances resolution along the axial direction [13, 19].

Just like the imaging PSF the illumination pattern will also be diffraction limited; its corresponding Fourier peaks will lie within the support of the systems PSF. The resolution improvement is thus limited to a factor of about two (slightly better, as the illumination support is greater than the detection support due to the Stokes shift of wavelengths in the fluorescence, or possible TIRF illumination). Exploiting non-linearities in the sample's response to the illumination light, higher harmonics can be generated in the effective illumination (4.1.3), with peaks lying outside the OTF support. This way SIM can in principle improve the resolution by an arbitrary amount.

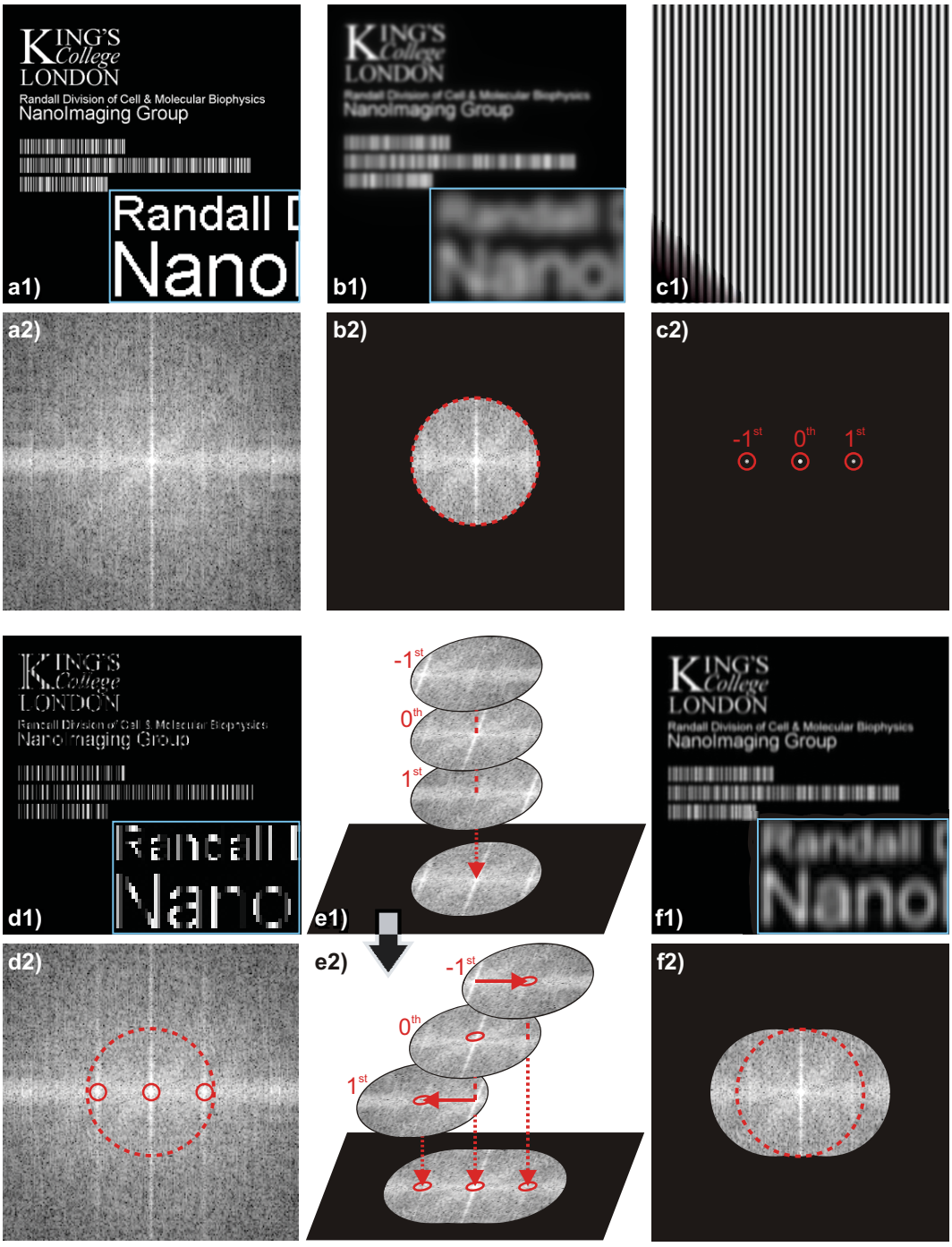


Fig. 3.3: Graphical illustration of the principle of high resolution SIM. For a detailed explanation see 3.2.2.

4. STRUCTURED ILLUMINATION IN PERFECT SYSTEMS

This section describes the technique of structured illumination for resolution improvement – the image formation and reconstruction algorithms – for perfect systems. In the presence of imperfections in the systems or unknown parameters, more sophisticated algorithms have to be employed, which will be described in chapter 5.

4.1 Image formation

4.1.1 Illumination

The grating used for illumination is projected into the sample through the objective and is therefore diffraction-limited in itself. It can be generated by any number of foci in the back focal plane of the objective, the different amplitude orders. Every such order generates a plane wave $E_a \exp\{\imath(\vec{k}'_a \cdot \vec{r} + \varphi_a)\}$ in sample space, propagating at an angle defined by the position of the corresponding BFP point's projection $\vec{k}'_a$ onto the McCutchen's generalised aperture [42] (see Ewald's sphere [12]). E_a are the electric field amplitudes and φ'_a the phases of the individual amplitude orders. The interference pattern between these waves is the intensity excitation pattern:

$$\begin{aligned} L(\vec{r}) &= \left(\sum_a E_a \exp\{\imath(\vec{k}'_a \cdot \vec{r} + \varphi'_a)\}\right)\left(\sum_a E_a \exp\{\imath(\vec{k}'_a \cdot \vec{r} + \varphi'_a)\}\right)^* \\ L(\vec{r}) &= \sum_n L_n \exp\{\imath(\vec{k}_n \cdot \vec{r} + \varphi_n)\}, \end{aligned} \tag{4.1}$$

where in the second step we have gone from an amplitude notation to one using intensity orders with individual strengths L_n.

In SIM the sample is usually illuminated with two (Fig. 4.1a, blue) or three (Fig. 4.1b,

blue) amplitude orders, which can be generated using a diffraction grating, or a spatial light modulator displaying a periodic structure. This two- or three-point illumination results in three (Fig. 4.1a, red) or seven (Fig. 4.1b, red) intensity orders, which can be readily seen when looking at Eq. 4.1 in Fourier space, where taking the absolute square of the plane waves corresponding to amplitude orders translates to an autocorrelation of the amplitude orders' Fourier peaks.

The resulting intensity orders have k-vectors corresponding to the possible differences of amplitude k-vectors. Because the acquisition of a two-dimensional image of the three-dimensional (3D) sample corresponds to a sum projection along $\vec{k}_z$ in Fourier space, orders with identical lateral frequency components $\vec{k}_{n,xy}$ can usually be treated as one order, so that for three amplitude orders, we end up with five intensity orders[1]. In order to do this, some orders need to be associated with modified PSFs. This is explained in more detail in 4.2.2.

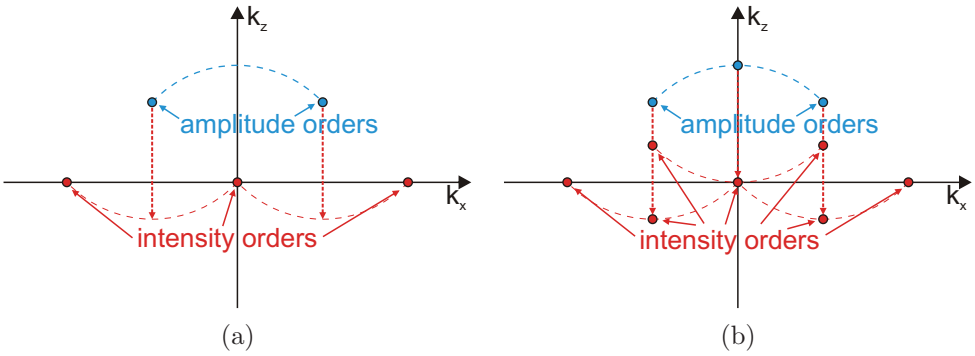


Fig. 4.1: Illumination patterns are normally generated through interference of plane waves, represented by points in Fourier space – the amplitude orders (blue). The intensity distribution of the illumination is represented by the intensity orders (red) and are obtained through an auto-correlation of the amplitude orders. For two amplitude orders this results in three intensity orders (a) whereas three amplitude orders yield seven intensity orders (b) in 3D Fourier space.

If the phases φ_n are modulated through a translation of the grating or the pattern

[1] This is not true, if the phases of orders with identical lateral frequency components are modulated independently. This can be the case if the amplitude orders are phase shifted individually, rather than in a correlated manner as is the case when phase-shifting is accomplished through the physical translation of a diffraction grating. If their phases are modulated independently, the orders will have to be treated individually.

displayed on a spatial light modulator – rather than individually modulating each amplitude order – their modulation will not be independent but rather depend on the grating (or pattern) position $\vec{s}$; Eq. 4.1 is changed to

$$L(\vec{r}) = \sum_n L_n \exp\{\imath(\vec{k}_n \cdot \vec{r} + \vec{k}_{n,xy} \cdot \vec{s})\}.$$

4.1.2 Fluorescence emission

The sample fluorophores will be excited by the illumination light and emit fluorescence. If the fluorophore response to the illumination is linear, the light emitted by the sample can be written as a simple product $S(\vec{r})L(\vec{r})$ of the sample's fluorophore density distribution S and the illumination pattern L[2].

If the sample response is not linear, the relation between sample and illumination will be more complex [25]. However, it can still be written as a product of the sample's fluorophore distribution $S(\vec{r})$ and a modified *effective* illumination $L_{\text{eff}}(\vec{r})$.

4.1.3 Nonlinear sample response: effective illumination

As was explained in 3.2.2, SIM improves the resolution by extending the support of the OTF by $\vec{k}_{n,\max}$ (the position of the highest order in frequency space). Because the illumination pattern itself is diffraction-limited, $\vec{k}_{n,\max}$ has to lie within the OTF support and thus the maximum gain in resolution possible with linear structured illumination is two-fold (slightly higher because of the shorter wavelength used in the illumination due to the Stokes shift).

However, a nonlinear response of the sample to the illumination light (e.g., saturating optical transitions in conventional [16] or preferably photo-switchable fluorophores [32]) can introduce higher harmonics so that the effective illumination pattern will no longer be diffraction limited (Fig. 4.2). This can be easily seen when looking at the nonlinear sample response for fluorescence saturation, which we write as the effective illumination [25] $L_{\text{eff}}(\vec{r}) = abL(\vec{r})/(b + aL(\vec{r}))$, where a is proportional to the

[2] For simplicity constant scaling factors such as quantum efficiency of the fluorescent dyes, etc. have been omitted, or are assumed to be included in S.

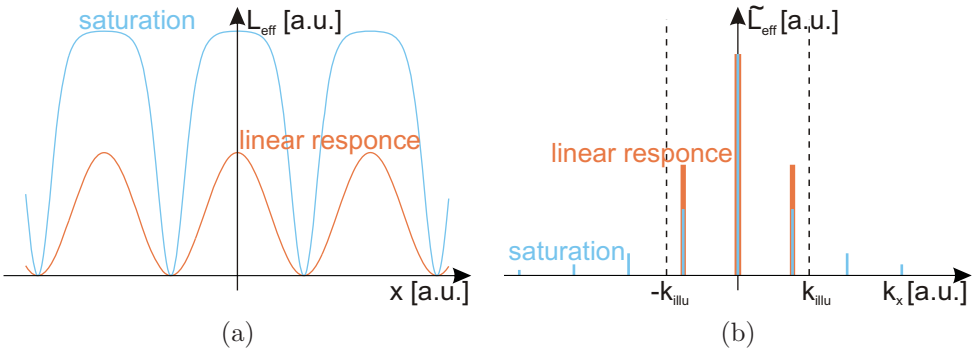


Fig. 4.2: Effective illumination under sinusoidal grating illumination. For a linear fluorescence response (orange) the effective illumination will also be sinusoidal (a), the Fourier transform will consist of three peaks (b). If the intensity is high enough for the fluorescence to become saturated (blue), the nonlinear response yields an effective illumination that is no longer sinusoidal (a). Its Fourier transform will contain higher harmonics, which may lie outside the support of the illumination OTF (b). Note that for better perceptibility the two curves in (a) are not scaled correctly; in order to achieve saturation the intensity would have to be much higher than for a linear response. Also in (b) the Fourier transforms are normalised to the same mean intensity.

fluorescence rate constant and b is inversely proportional to the absorption cross section and the lifetime of the fluorophores (see appendix A.2 for a derivation). For low illumination intensities L or large b (i.e. short lifetime or small absorption cross section) the relationship between effective and true illumination is linear. For high intensities L or small b fluorescence becomes saturated and reaches a plateau.

A series expansion around $L(\vec{r})/L_0 = 0$ yields $L_{\text{eff}}(\vec{r}) = -b\sum_{j=1}^{\infty}(-aL(\vec{r})/b)^j$, where the presence of higher harmonics is obvious in the exponent terms $L^j(\vec{r})$.

This effective nonlinear illumination can still be written as Eq. 4.1, however with an increased number of intensity orders N. As the effective illumination equals the true illumination in the case of a linear sample response, it will hereafter be referred to as L rather than L_{eff} for reasons of simplicity.

4.1.4 Image acquisition

The fluorescence light emitted by the sample is then captured by the microscope objective and generates an intensity distribution in image space, which can be described

by a convolution of the emitted light with the system's PSF[3]:

$$I(\vec{r}) = \left[(S^{(z_0)} L) \otimes h \right] (\vec{r}),$$
$$S^{(z_0)}(\vec{r}) := S(x, y, z - z_0).$$

Here refocussing to a plane $z = z_0$ in the sample is achieved by physically moving the sample by a distance z_0 along the axial direction (denoted by $S^{(z_0)}$), while the illumination pattern remains stationary with respect to the objective.

The resulting light distribution $I(\vec{r})$ is then sampled by a camera placed in the focal plane of image space, yielding a two-dimensional image

$$I^{(z_0)}(\vec{r}_{xy}) = \left[(S^{(z_0)} L) \otimes h \right] (\vec{r}) \Big|_{z=0}.$$

In order to be able to separate the information components attached to the individual orders computationally, we have to take at least as many images as there are independent orders (i.e. N), while varying the illumination pattern's position (also called grating phase) from image to image[4]. This has to be done in a way that the grating phase is different in at least N of the images – or more accurately, at least N of the images have to be linearly independent of each other. The grating phase can be incorporated into Eq. 4.1 as an additional phase which applies differently to different orders, depending on their lateral frequency position. As a (lateral) grating shift by $\vec{s}_m$ in the m^{th} image corresponds to a multiplication of the grating's Fourier pattern with $\exp\{\imath \vec{k} \cdot \vec{s}_m\}$, the n^{th} order's peak at $\vec{k}_n$ will acquire an additional phase of $\exp\{\imath \vec{k}_n \cdot \vec{s}_m\}$ in the wave term. Together with Eq. 4.1 we get for the m^{th} image

$$I_m^{(z_0)}(\vec{r}_{xy}) = [S^{(z_0)}(\vec{r}) \sum_n L_n \exp\{\imath(\vec{k}_n \vec{r} + \vec{k}_n \cdot \vec{s}_m)\}] \otimes h(\vec{r})|_{z=0},$$

[3] Constant scaling factors such as detector efficiency have been omitted for simplicity, or are assumed to be included in the PSF.

[4] This is only true if the information is to be separated using a straightforward inversion of the equation system. When using the fact that the same information is present more than once in the Fourier images, this knowledge can be used to separate the information with fewer images [20].

which in Fourier space transforms to

$$\tilde{I}_m^{(z_0)}(\vec{k}_{xy}) = \int_{-\infty}^{\infty} \sum_n L_n \exp\{\imath(\vec{k}_n \cdot \vec{s}_m)\}[(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_n)]\tilde{h}(\vec{k})\, dk_z.$$

When we see $\tilde{I}_m^{(z_0)}(\vec{k}_{xy})$ as the m^{th} element of a vector of images $\vec{\tilde{I}}^{(z_0)}(\vec{k}_{xy})$, we can simplify the above equation using a matrix multiplication:

$$\vec{\tilde{I}}^{(z_0)}(\vec{k}_{xy}) = \bar{M}\vec{\tilde{O}}^{(z_0)}(\vec{k}_{xy}), \tag{4.2}$$

where $\bar{M}$ is the mixing matrix

$$\bar{M}_{mn} = \exp\{\imath(\vec{k}_n \cdot \vec{s}_m)\}$$

and $\vec{\tilde{O}}^{(z_0)}(\vec{k}_{xy})$ is a vector containing the individual "components", which are defined as

$$\tilde{O}_n^{(z_0)}(\vec{k}_{xy}) = L_n \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_n)]\tilde{h}(\vec{k})\, dk_z. \tag{4.3}$$

4.2 Image reconstruction

Although the OTF limits the spatial frequencies that can be captured by the imaging system, the shift in sample information $\tilde{S}(\vec{k}-\vec{k}_n)$ transposes high frequency information to lower frequencies so that this information will now lie within the support region of the OTF and thus be captured. If this information can be isolated (separated) and computationally shifted back to its original position in frequency space, an image with an extended frequency support and thus a higher resolution can be reconstructed.

4.2.1 Separation of the superimposed components

According to Eq. 4.2 and Eq. 4.3, the different Fourier images $\tilde{I}_m^{(z_0)}(\vec{k}_{xy})$ contain information from several regions in Fourier space in the form of superimposed components

$\tilde{O}_n^{(z_0)}(\vec{k}_{xy})$. In order to separate these components we invert Eq. 4.2:

$$\vec{\tilde{O}}^{(z_0)}(\vec{k}_{xy}) = \bar{M}^{-1}\vec{\tilde{I}}^{(z_0)}(\vec{k}_{xy}). \tag{4.4}$$

If the mixing matrix $\bar{M}$ is not square (i.e. more images were taken then there are orders to separate), its inverse $\bar{M}^{-1}$ can be calculated as the Penrose pseudo inverse [47].

The fact that $\bar{M}$ has to be reversible imposes certain requirements for the acquisition of images. Firstly, the number of images has to be at least as high as the number of components to be separated. Secondly, the grating positions $\vec{s}_m$ have to be chosen in such a way, that all of $\bar{M}$'s column vectors will be linearly independent, i.e. the matrix may not be singular. The first requirement follows from the stronger second requirement.

4.2.2 Combining components of identical lateral shift

As the different orders' phase modulations with grating movement ($\exp\{\imath\vec{k}_n \cdot \vec{s}_m\}$) depend on the order peaks' lateral frequency component, any components attached to orders with the same lateral frequency component will be modulated the same way and therefore cannot be separated from one another through the unmixing process of Eq. 4.4. Instead, all such inseparable components can be – or rather have to be – treated as one component.

This leads us to a new definition of components (see appendix A.3)

$$\tilde{\Omega}_n^{(z_0)}(\vec{k}_{xy}) = \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - n\vec{k}_\mathrm{g})]\tilde{h}_n(\vec{k})\,dk_z,$$

with order-dependent OTFs

$$\tilde{h}_n(\vec{k}) = \sum_{i\in\mathbb{I}} L_i \tilde{h}(\vec{k} + \vec{k}_{i,z}),\ \mathbb{I} = \{i :\ \vec{k}_{i,xy} = n\vec{k}_\mathrm{g}\}. \tag{4.5}$$

Here we have $-(N-1)/2 \leq n \leq (N-1)/2$, where N is now the number of combined orders (i.e. the number of orders with independent lateral frequency positions). The

mixing matrix $\bar{M}$ has to be changed accordingly (i.e. recurring columns have to be removed):

$$\bar{M}_{mn} = \exp\{\imath(n\vec{k}_\mathrm{g} \cdot \vec{s}_m)\} \tag{4.6}$$

This has the benefit of reducing the number of independent components, which means that fewer images are required for their separation.

If the grating is stepped evenly ($|\vec{s}_{m+1} - \vec{s}_m| = 2\pi/(M|\vec{k}_\mathrm{g}|)$), the component unmixing process can be easily calculated as a Fourier transform of the series of input images $\tilde{I}_m^{(z_0)}$ with respect to the coordinate of image number m.

4.2.3 Shifting the sample information

Once the individual components $\tilde{\Omega}_n^{(z_0)}$ have been separated they can be shifted by $-n\vec{k}_\mathrm{g}$ to their true origins in Fourier space. Because of the sum projection along k_z (Eq. 4.3) it is enough to shift the components laterally; in fact, performing this operation on a two-dimensional image leaves no other possibility.

These shifted components are

$$\tilde{\Omega}_n^{(z_0)}(\vec{k}_{xy} + n\vec{k}_\mathrm{g}) = \int_{-\infty}^{\infty} [\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})]\tilde{h}_n(\vec{k} + n\vec{k}_\mathrm{g})\,dk_z,$$

with the accordingly shifted order-dependent OTFs

$$\tilde{h}_n(\vec{k} + n\vec{k}_\mathrm{g}) = \sum_{i\in\mathbb{I}} L_i\tilde{h}(\vec{k} + \vec{k}_i),\ \mathbb{I} = \{i :\ \vec{k}_{i,xy} = n\vec{k}_\mathrm{g}\}.$$

These components contain sample information from different frequency bands, defined by the shifted OTFs. Combining these shifted components yields sample information from an effective OTF with extended support, as illustrated in Fig. 4.3.

4.2.4 Transition from two- to three-dimensional components

If a whole z-stack of images is acquired for evenly spaced z_0, a Fourier transform with respect to z_0 yields the complete three-dimensional information of the sample in the

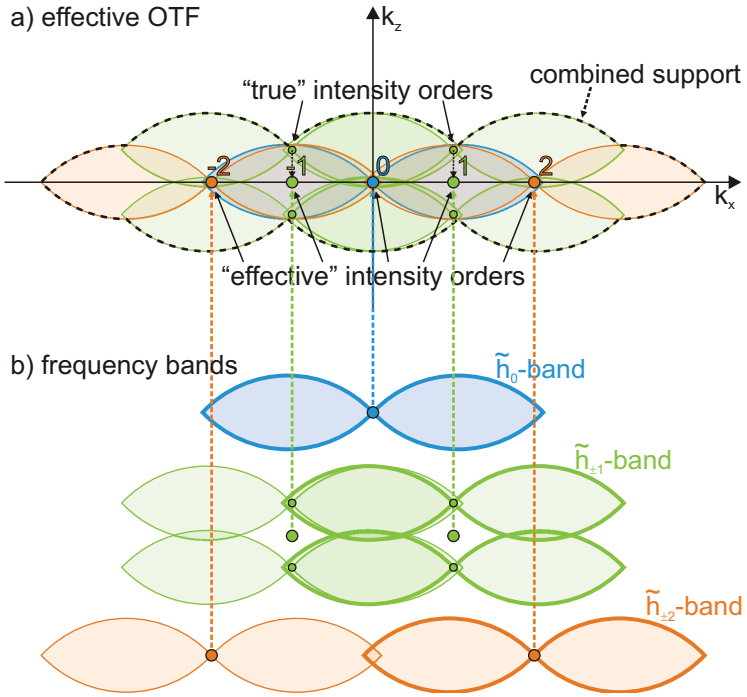


Fig. 4.3: (a) Extended OTF support. Once the information contained in the individual shifted components has been separated and shifted to its true origin in Fourier space, the components can be recombined. This yields a reconstructed image with information from within an effective OTF with increased frequency support (black dotted line). (b) Different components correspond to different frequency bands, shown in different colours, while the effective OTFs $\tilde{h}_1, \tilde{h}_2, \tilde{h}_3$ are shown with a bold outline. In nonlinear SIM the number of bands will be greater than three and each of the effective orders may contain more than just two true intensity orders.

form of three-dimensional components[5] (see appendix A.4):

$$\tilde{\Omega}_n(\vec{k} + n\vec{k}_g) = \tilde{S}(\vec{k})\tilde{h}_n(\vec{k} + n\vec{k}_g).$$

4.2.5 Recombination of shifted components

The different frequency bands contain sample information from within an extended support (Fig. 4.3). A recombination of all components will yield a Fourier image $\tilde{J}(\vec{k})$

[5] In a perfect system the transition from 2D to 3D could be made earlier. However, imperfect systems may require the mixing matrix $\bar{M}$ to be adjusted for each focal position z_0 (e.g. if the grating movement is not exactly reproducible). In this case the components have to be separated slice by slice before they can be combined into 3D components.

with a new effective OTF.

Summation of components

There are several ways of combining the separated components. The simplest recombination is a summation of shifted components [22]:

$$\begin{aligned}\tilde{J}_{\mathrm{sum}}(\vec{k}) &= \sum_n \tilde{\Omega}_n(\vec{k}+n\vec{k}_{\mathrm{g}}) \\ &= \tilde{S}(\vec{k})\sum_n \tilde{h}_n(\vec{k}+n\vec{k}_{\mathrm{g}}) \\ &= \tilde{S}(\vec{k})\tilde{h}_{\mathrm{sum}}(\vec{k}).\end{aligned}$$

This corresponds to imaging the sample with an effective OTF

$$\tilde{h}_{\mathrm{sum}}(\vec{k}) = \sum_n \tilde{h}_n(\vec{k}+n\vec{k}_{\mathrm{g}}). \tag{4.7}$$

Weighted averaging in Fourier space

While the summation of all individual components does yield a reconstructed image $\tilde{J}_{\mathrm{sum}}$ with an improved effective OTF $\tilde{h}_{\mathrm{sum}}$, in light of finite signal-to-noise levels this recombination is not optimal. Assuming a constant mean intensity in all acquired images, all components will contain white noise of the same mean standard deviation σ, whereas the level of information content at a certain frequency for a shifted component $\tilde{\Omega}_n(\vec{k}+n\vec{k}_{\mathrm{g}})$ depends on the magnitude of the corresponding shifted OTF $\tilde{h}_n(\vec{k}+n\vec{k}_{\mathrm{g}})$. By adding information of a component with very little (or even no) transfer strength to information with strong transfer strength, one adds little (or no) information, while the noise level is in fact raised. In this case it would be better to use mainly (or only) the information of the component for which the OTF has a strong transfer strength.

Regarding SNR the ideal recombination uses weighted averaging in Fourier space [26] (see appendix A.5):

$$\tilde{J}_{\mathrm{wa}}(\vec{k}) = \frac{\sum_n \tilde{w}_n(\vec{k})\tilde{\Omega}_n(\vec{k}+n\vec{k}_{\mathrm{g}})}{\sum_n \tilde{w}_n(\vec{k})}, \tag{4.8}$$

a summation of components with order and frequency dependent weights

$$\tilde{w}_n(\vec{k}) = \tilde{h}_n(\vec{k} + n\vec{k}_\mathrm{g})/\sigma_n^2, \ \forall n. \tag{4.9}$$

The reconstructed Fourier image is

$$\begin{aligned}\tilde{J}_\mathrm{wa}(\vec{k}) &= \tilde{S}(\vec{k})\tilde{h}_\mathrm{wa}(\vec{k}),\\ \tilde{h}_\mathrm{wa}(\vec{k}) &= \sum_n \{\tilde{h}_n^2(\vec{k} + n\vec{k}_\mathrm{g})/\sigma_n^2\} / \sum_n \{\tilde{h}_n(\vec{k} + n\vec{k}_\mathrm{g})/\sigma_n^2\}\end{aligned}$$

and has a standard deviation of

$$\sigma_\mathrm{wa}(\vec{k}) = \frac{\sqrt{\sum_n \tilde{w}_n^2(\vec{k})\sigma_n^2}}{\sum_n \tilde{w}_n(\vec{k})}, \tag{4.10}$$

where σ_n^2 is the variance of the noise of the individual components. Here we allow the noise level to differ from component to component, but assume white, frequency independent noise[6].

4.2.6 Normalising the noise

The resulting noise distribution (Eq. 4.10) is no longer frequency independent. For a fair comparison of SIM's imaging capabilities with other techniques one should look at an image that has been corrected to once again have white noise. For this we need to divide the Fourier image by the standard deviation of its own noise component. The result

$$\begin{aligned}\tilde{J}_\mathrm{wn}(\vec{k}) &= \tilde{J}_\mathrm{wa}(\vec{k})/\sigma_\mathrm{wa}(\vec{k})\\ &= \sum_n \{\tilde{w}_n(\vec{k})\tilde{\Omega}_n(\vec{k} + n\vec{k}_\mathrm{g})\} / \sqrt{\sum_n \tilde{w}_n^2(\vec{k})\sigma_n^2}\\ &= \sum_n \{\tilde{h}_n(\vec{k})\tilde{\Omega}_n(\vec{k} + n\vec{k}_\mathrm{g})/\sigma_n^2\} / \sqrt{\sum_n \tilde{h}_n^2(\vec{k} + n\vec{k}_\mathrm{g})/\sigma_n^2}\end{aligned} \tag{4.11}$$

[6] Frequency dependent noise can be easily accounted for by making the variance term frequency dependent, i.e. $\sigma_n^2(\vec{k})$.

is a Fourier image with an effective OTF that contains white noise at optimum signal-to-noise ratio. This effective imaging OTF is

$$\tilde{h}_{\mathrm{wn}}(\vec{k}) = \sqrt{\sum_n \{\tilde{h}_n^2(\vec{k} + n\vec{k}_{\mathrm{g}})/\sigma_n^2\}}. \tag{4.12}$$

Figure 4.4 compares this effective imaging OTF $\tilde{h}_{\mathrm{wn}}$ (blue) to that obtained through a mere summation of components, $\tilde{h}_{\mathrm{sum}}$ (Eq. 4.7, red). The curves show cross-sections through OTFs, corresponding to SIM imaging for three directions and illumination with three equal amplitude orders, resulting in five intensity orders with relative strengths $\{1, 2, 3, 2, 1\}$. Furthermore the figure shows a cross-section through the conventional wide-field OTF, $h_{\text{wide-field}}$ (green). Note that all curves have been normalised to have the same white noise, allowing a fair comparison. Both SIM OTFs transmit much higher frequencies than the wide-field OTF, which has no support for frequencies greater than k_{max}. The central value of the white-noise, weighted-averaging OTF $\tilde{h}_{\mathrm{wn}}$ is approximately 50% higher than that of the summation OTF $\tilde{h}_{\mathrm{sum}}$, leading to higher signal levels in the weighted-averaging image. Furthermore, $\tilde{h}_{\mathrm{wn}}$ transmits high frequencies significantly better than $\tilde{h}_{\mathrm{sum}}$, resulting in better resolution images.

For frequencies $\vec{k}_{\mathrm{out}}$ outside the support of any of the shifted OTFs $\tilde{h}_n(\vec{k} + n\vec{k}_{\mathrm{g}})$, all weights will be zero, in which case Eq. 4.8 and Eq. 4.10 are not defined. For these frequencies we can set the Fourier image to $\tilde{J}(\vec{k}) = 0, \forall \vec{k} = \vec{k}_{\mathrm{out}}$, as here it will contain no sample information. However, this would lead to strongly non-white noise in the reconstructed image. Whereas true white noise has no inter-pixel correlation between pixels in the real space image, limiting the white noise to only the region of support of the effective OTF, leads to a convolution of this uncorrelated noise with a noise PSF. This noise PSF corresponds to the an inverse Fourier transform of the footprint[7] $\tilde{f}(\tilde{h}_{\mathrm{wn}}(\vec{k}))$ of the effective OTF. A thus shaped noise PSF cannot necessarily easily be discerned as noise by eye; the shot noise may therefore appear as image artefacts.

A better option may therefore be to fill up these empty Fourier regions ($\vec{k}_{\mathrm{out}}$) with

[7] The footprint $\tilde{f}$ of an OTF is 1 where the OTF has support and 0 otherwise.

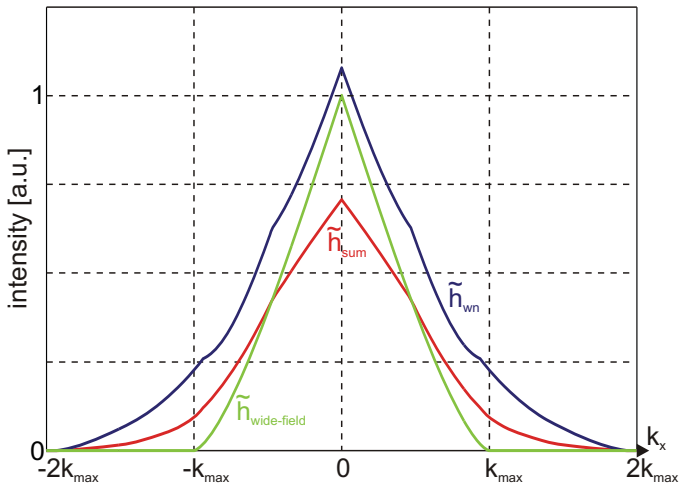


Fig. 4.4: Comparison of summing, weighted-averaging and wide-field OTFs. The (white-noise normalised) weighted-averaging OTF (blue) has a 50% higher central value then the OTF obtained through a mere summation of components (red), resulting in higher signal levels. Furthermore it transmits high frequencies significantly better, leading to higher resolution images. Unlike the wide-field OTF (green), both SIM OTFs transmit frequencies outside the wide-field support, which is limited to $k_{\max}$.

white random noise of the same level as inside the support region of the OTF.

This, however leads to the "additional" noise being completely independent of intensities in the reconstructed real space image $J(\vec{r})$. As a result, previously dark areas in the image may now also exhibit an added noise component.

To avoid this, one can fill the empty regions in the Fourier image with noise which actually corresponds to the real space image. To do so, Poisson noise is computationally added to the reconstructed real space image. The Fourier transform of this noise image will exhibit white, but intensity correlated noise outside the effective OTF support. This noise can now be added to the areas of no support of the original (non-noise added) reconstructed Fourier image.

4.2.7 Wiener filter deconvolution

Although the white noise image may most closely resemble a "natural" image, it is beneficial to further improve the image through deconvolution. In the imaging process some frequencies are transmitted better than others. Deconvolution tries to reverse this

effect by enhancing weaker frequencies and to thus achieve a more true representation of sample information. However, simply reversing the effect of the convolution through a convolution with the inverse effective PSF (or multiplication with the inverse effective OTF in Fourier space) would not just enhance the dampened sample information but also unduly magnify the noise contributions for frequencies of weak signal-to-noise levels.

Wiener Filter deconvolution attempts to minimise the impact of deconvolved noise. It does so by minimising the mean square error (MSE) of the deconvolved Fourier image.

The reconstructed white noise image is $\tilde{J}_{\mathrm{wn}}(\vec{k}) = \tilde{S}(\vec{k})\tilde{h}_{\mathrm{wn}}(\vec{k}) + \tilde{n}(\vec{k})$, where $\tilde{n}(\vec{k})$ is white noise with a standard deviation σ_{wn}. Our goal is to find a Wiener filter function $\tilde{W}(\vec{k})$ so that the filtered image

$$\tilde{J}_{\mathrm{wien}}(\vec{k}) = \tilde{J}_{\mathrm{wn}}(\vec{k})\tilde{W}(\vec{k}) \tag{4.13}$$

has a minimal MSE $\varepsilon(\vec{k}) = \langle|\tilde{S}(\vec{k}) - \tilde{J}_{\mathrm{wien}}(\vec{k})|^2\rangle$ as compared to the true sample information (see appendix A.6).

If we assume the imaging PSF to be real and symmetric, so will be the OTF $\tilde{h}_{\mathrm{wn}}(\vec{k})$ and we can thus assume the Wiener function $\tilde{W}(\vec{k})$ to be real and symmetric as well. This assumption yields the Wiener filter function

$$\tilde{W}(\vec{k}) = \frac{\tilde{h}_{wn}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle}{\tilde{h}^2_{\mathrm{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle + \langle|\tilde{n}(\vec{k})|^2\rangle}.$$

This Wiener filter deconvolution can be incorporated directly into the equation for order recombination (see Eq. 4.11 and Eq. 4.13). This yields

$$\begin{aligned}\tilde{J}_{\mathrm{wien}}(\vec{k}) &= \frac{\langle|\tilde{S}(\vec{k})|^2\rangle \sum_n \{\tilde{h}_n(\vec{k}+n\vec{k}_{\mathrm{g}})\tilde{\Omega}_n(\vec{k}+n\vec{k}_{\mathrm{g}})/\sigma_n^2\}}{\langle|\tilde{S}(\vec{k})|^2\rangle \sum_n \{\tilde{h}_n^2(\vec{k}+n\vec{k}_{\mathrm{g}})/\sigma_n^2\} + \langle|\tilde{n}(\vec{k})|^2\rangle} \\ &= \frac{\langle|\tilde{S}(\vec{k})|^2\rangle \tilde{h}^2_{\mathrm{wn}}(\vec{k})\tilde{S}(\vec{k})}{\langle|\tilde{S}(\vec{k})|^2\rangle \tilde{h}^2_{\mathrm{wn}}(\vec{k}) + \langle|\tilde{n}(\vec{k})|^2\rangle}.\end{aligned} \tag{4.14}$$

The white noise variance $\langle|\tilde{n}(\vec{k})|^2\rangle$ is constant. As the sample's power spectrum $\langle|\tilde{S}(\vec{k})|^2\rangle$ is unknown, this value is usually replaced with a constant. This is an approach we follow in our reconstruction algorithm, where the constant has to be adjusted somewhat empirically from experiment to experiment. The same approach was used for 3D SIM by Gustafsson et al. [19].

Deficiencies of Wiener filter deconvolution

Wiener filter deconvolution is achieved by multiplication with a Wiener filter function in Fourier space. In real space this corresponds to a convolution. The function used for this convolution is the same over the whole image to be deconvolved. It does not depend on local intensities and therefore neither on local signal-to-noise levels (considering photon noise, SNR levels depend strongly on local image intensity). It therefore cannot be an ideal deconvolution but rather compromises between optimising different local signal-to-noise levels.

4.2.8 Apodisation function

The Wiener filter deconvolution is not ideal, as it makes the invalid assumption of a constant sample power spectrum. As "natural" samples tend to have higher power in lower frequencies, the assumption of a constant sample spectrum puts undue emphasis on high frequencies. As it is the high frequencies that have the lowest SNR, their noise will be amplified too much. This can be avoided by multiplying the reconstructed Fourier image with an apodisation function $\tilde{a}(\vec{k})$ dampening higher frequencies:

$$\tilde{J}_{\text{final}}(\vec{k}) = \tilde{J}_{\text{wien}}(\vec{k})\tilde{a}(\vec{k}).$$

There are several choices for the apodisation function. Gustafsson et al. [19] use a three-dimensional triangular function decreasing linearly from unity in the centre to zero at the border of the 3D support of the estimated sample.

In our algorithms we use two different approaches, both of which are based on the

Euclidean distance transform[8] d of the OTF's footprint $\tilde{f}$.

Based on this function we define our first apodisation function as

$$\tilde{a}_1(\vec{k}) := [d(\tilde{f}(\tilde{h}(\vec{k})))]^{\kappa}.$$

In this definition the parameter $\kappa \geq 0$ lets us adjust the shape of the apodisation function, as shown in Fig. 4.5. For $\kappa = 1$ we get the unchanged distance transform and thus a linear apodisation function, somewhat similar to that of Gustafsson et al. [19]. A choice of $\kappa \geq 1$ curves the previously linear apodisation function to make it concave, emphasising lower frequency information, whereas $\kappa \leq 1$ yields a convex apodisation function emphasising higher frequency information.

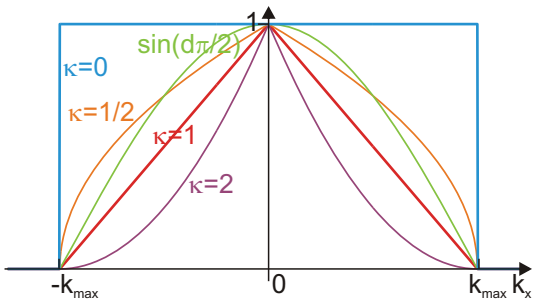


Fig. 4.5: Apodisation functions for weighted averaging, shown in a 1D example. The blue rectangular function represents the footprint of the effective OTF. It is 1 inside the support and 0 otherwise. The normalised distance transform of the footprint (red, $\kappa = 1$) is defined as the distance to the edge of the support. Taken to the power of a factor κ this yields the apodisation function used for apodising the sample information. For $0 < \kappa < 1$ (e.g. $\kappa = 1/2$, orange) the apodisation function becomes convex, emphasising higher spatial frequencies. For $\kappa > 1$ the apodisation function will be concave, emphasising lower spatial frequencies (e.g. $\kappa = 2$, purple). Another type of apodisation function ($\sin(d\pi/2)$, green) apodises the data in a sinusoidal way (see Eq. 4.15).

The second type of apodisation function apodises the data in a sinusoidal fashion [25] and is defined as

$$\tilde{a}_2(\vec{k}) := \sin(\frac{\pi}{2} d(\tilde{f}(\tilde{h}(\vec{k})))). \tag{4.15}$$

[8] The Euclidean distance transform $d(\tilde{f}(\vec{k}))$ is defined as the closest distance of a point $\vec{k}$ to the edge of $\tilde{f}$, normalised so that the maximum distance (usually at the centre) is 1.

This type of filter is also known as “sine window”.

4.3 Example

Figure 4.6 shows an example of SIM image reconstruction using the algorithm described above. The simulation parameters can be found in appendix A.7. Figures 4.6a, d show the sample (a photograph of London’s King’s cross railway station) and its Fourier transform. Figure 4.6b shows the simulated wide-field image. Its Fourier transform (Fig. 4.6e) contains no information outside the support of the wide-field OTF (marked by the blue circle), resulting in a loss in resolution as compared to the original sample. SIM extends the effective support (Fig. 4.6f, orange boundary), recovering a lot of image detail (Fig. 4.6c) lost in the conventional image.

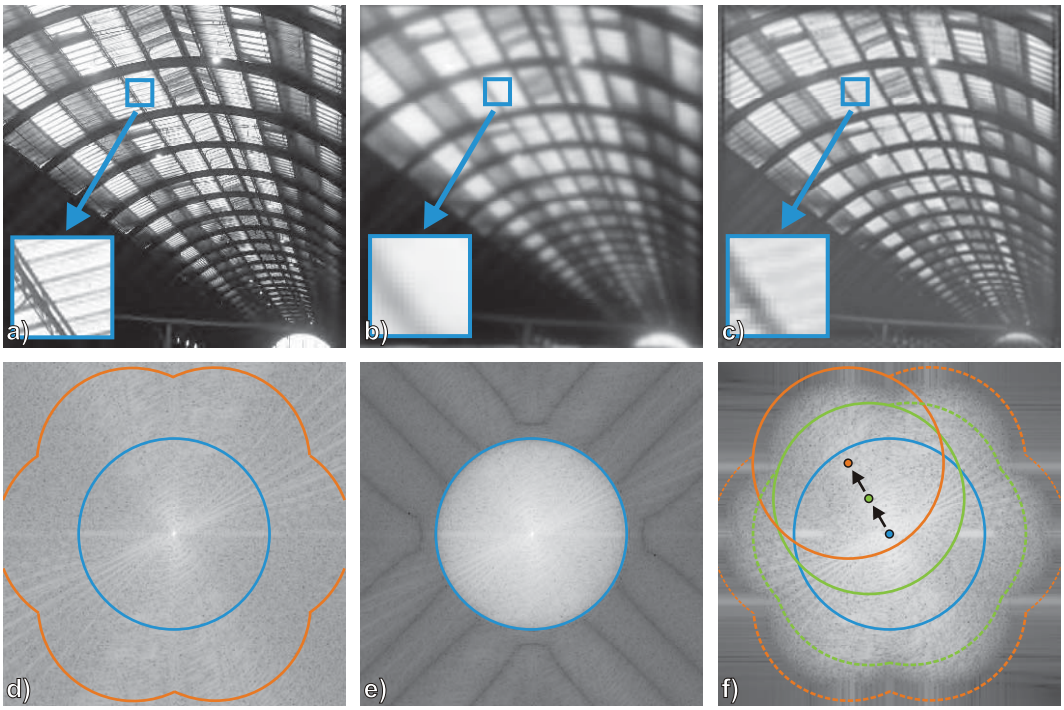


Fig. 4.6: Simulation of SIM image reconstruction using the described algorithm. (a) shows the synthetic sample used for the simulation, (d) its Fourier transform. (b) shows the simulated wide-field image. Its Fourier transform (e) contains no information outside the support of the wide-field OTF (marked by the blue circle), resulting in a loss in resolution as compared to the original sample. (f) SIM extends the effective OTF support (orange boundary), recovering a lot of image detail (c) lost in the conventional image. Information toward the edge of the support was dimmed to avoid noise artefacts. The different colours in the Fourier image (f) represent different frequency bands, as shown in Fig. 4.3. Real space images are 12.8 µm wide.

5. IMAGE RECONSTRUCTION IN IMPERFECT SYSTEMS

If all experimental parameters are known with high precision the reconstruction of SIM images as described in chapter 4 is relatively straightforward. In reality the experimental setups often cannot guarantee precise enough knowledge or control of all parameters for an artefact free reconstruction of the data, warranting the algorithmic determination of these unknown parameters. Several of these parameters can in principle be precisely controlled using sophisticated setups [19]; however even the most perfect systems will still require certain parameters to be fitted.

Our approach to reconstruction of structured illumination data was to assume imperfect setups and try to algorithmically correct for these imperfections. While it is possible to construct setups not requiring such sophisticated reconstruction algorithms, we believe that such algorithms are beneficial to the scientific community, as they greatly simplify the efforts of constructing functional SIM setups.

In this section I describe what potential problems may arise in the reconstruction of SIM data, what their typical effects are on image reconstruction, and what measures our algorithms take to correct for these unknowns. These algorithms were developed together with Rainer Heintzmann and Ondrej Mandula and have in part been presented in Mr. Mandula's master's thesis [41].

5.1 Camera background

Besides the actual image information the raw CCD images $I_m^{(z_0)}(\vec{r}_{xy})$ also contain an offset signal and thermal noise. In the absence of other experimental imperfections this background signal does not influence the actual reconstruction: after component-separation it will only be present in the zero order component and would therefore

simply contribute a background signal in the final image.

However, in combination with other experimental imperfections requiring adjusted reconstruction parameters (particularly the problems addressed in sections 5.3, 5.4, 5.6 and 5.7), the background information may also be present in the non-zero order components after separation. After shifting these components to their correct location in Fourier space, this residual background information will be in an incorrect location and cause artefacts in the final image.

5.1.1 Typical artefacts

In a first approximation, camera background can be seen as more or less uniform over the whole field of view; in Fourier space it therefore represents a sharp peak at the zero frequency. If this peak cannot be removed from all components, it will be present at multiples of the grating vector $\vec{k}_{\mathrm{g}}$ in the final Fourier image. This leads to periodic pattern artefacts in the final real space image. In the case of three grating directions, these peaks will form a honeycomb pattern similar to but – because the peaks are very narrow – even more pronounced than that visible in Fig. 5.3b.

5.1.2 Remedy: subtraction of background

To avoid these artefacts, a dark exposure image $I_{\mathrm{dark}}(\vec{r}_{xy})$ should be acquired. Ideally this is an average of ten or more single dark images[1]. This dark image contains the bias and thermal background and can be subtracted from each acquired SIM image to eliminate the unwanted background components: $I_m(\vec{r}_{xy}) \rightarrow I_m(\vec{r}_{xy}) - I_{\mathrm{dark}}(\vec{r}_{xy})$.

Alternatively, if the background is assumed to be constant over the whole field of view, it can be estimated from dark regions in the individual images.

Background subtraction is a standard image correction procedure used in most digital imaging applications.

[1] Be aware of possible of potential changes in background levels at the start of each acquisition series.

5.2 Non-uniform detector sensitivity and similar optical distortions

Images may further be distorted by varying detection sensitivity of the camera's pixel detectors. A similar distortion may also occur through dust or other contamination which can locally attenuate the detected signal.

5.2.1 Typical artefacts

The above distortions can be described as a multiplication of the true signal with a system dependent attenuation function $f_{\mathrm{att}}(\vec{r}_{xy})$. In Fourier space this corresponds to a convolution with $\tilde{f}_{\mathrm{att}}(\vec{k}_{xy})$. As the different components $\tilde{\Omega}_n(\vec{k})$ present in the Fourier images contain the information of the sample, $\tilde{S}(\vec{k}+n\vec{k}_{\mathrm{g}})$ shifted to different locations, the sample information $\tilde{S}(\vec{k})$ contained in the shifted components will be distorted differently from component to component.

This imperfect concurrence of information in overlapping shifted components may lead to insufficient optimisation of parameters where the optimisation depends on cross-correlations of separated components (sections 5.5, 5.6, 5.7, 5.9, 5.11).

5.2.2 Remedy: flat-field correction

To avoid these artefacts, a flat-field correction has to be performed. To do this, the attenuation $f_{\mathrm{att}}(\vec{r}_{xy})$ is determined from the acquisition of a pattern-free, full-field illumination image of a homogenous fluorescent sample (e.g. by imaging an approximately 10 µm thick fluorescent sea). The acquired SIM images are then flat-field corrected through a division with the attenuation function: $I_m(\vec{r}_{xy}) \rightarrow I_m(\vec{r}_{xy})/f_{\mathrm{att}}(\vec{r}_{xy})$.

Like background subtraction, flat-field correction is also a standard calibration procedure in most digital imaging applications.

5.3 Intensity fluctuations in the illumination

Intensity fluctuations between images stemming e.g. from fluctuations in the laser intensity will also lead to an imperfect separation of components. To correctly describe

the image formation under intensity fluctuations, every row vector of the mixing matrix $\bar{M}$ has to be multiplied with a factor proportional to the corresponding image's illumination intensity. Unmixing with the uncorrected inverse mixing matrix $\bar{M}^{-1}$ therefore will not yield correctly separated orders.

5.3.1 Typical artefacts

Incorrectly separated components are the most common cause for artefacts in the reconstructed images and can be effected by insufficient knowledge of many experimental parameters besides intensity fluctuations. Typical artefacts are false periodic patterns as can be seen in Fig. 5.3b. These artefacts are explained in more detail in 5.6, where an optimisation of the mixing matrix $\bar{M}$ is discussed.

5.3.2 Remedy: normalisation of image intensity

As the fluctuations in illumination intensity are usually not known, they cannot be accounted for in the mixing matrix $\bar{M}$. Instead, the individual images can be normalised to all contain the same integrated intensity, by dividing each image by its integrated intensity: $I_m^{(z_0)}(\vec{r}_{xy}) \rightarrow I_m^{(z_0)}(\vec{r}_{xy}) / \int I_m^{(z_0)}(\vec{r}'_{xy}) d^2r'$. Although they may now have different photon noise factors, the information contained in the images now corresponds to illumination under constant total intensity, so that the uncorrected inverse mixing matrix $\bar{M}^{-1}$ can be used for correct order separation.

If the camera background is not removed prior to this step, its contribution will now vary from image to image. Component-unmixing would therefore no longer remove the background from the higher orders components, warranting the subtraction of background before this normalisation.

However, this simple normalisation does not consider that some of the fluctuations in total intensity between images may not stem from fluctuations in illumination intensity, but may rather be the result of structured illumination of an inhomogeneous sample. This can readily be seen when considering a sparse bead sample: while for a certain grating position most beads may be illuminated, there may be other grating positions

for which those beads remain mostly dark. This leads to fluctuations in image intensity which are not the result of fluctuating illumination intensity and should in fact *not* be normalised. Though less dominant, the same effect can also be observed for less sparse biological samples as well.

Performance

Although it cannot account for fluctuations caused by sample inhomogeneities, normalisation is a good approximation: the simulation for random intensity fluctuations with a nominal variance of $(15\%)^2 = 0.0225$ resulted in actual fluctuations with a variance $(15.5\%)^2 = 0.0241$. After intensity normalisation these fluctuations were reduced to show a variance of $(0.32\%)^2 = 10^{-5}$. Compared to some biological samples, the synthetic sample used in the simulation is relatively dense. For sparser samples the normalisation may not yield as good results, but empiric observations nevertheless show good reconstruction results for sparse samples, as well.

5.4 Sample drift between individual images

If the sample experiences thermal lateral drift, and if this drift $\Delta\vec{r}_m$ is fast enough to be noticeable between individual images $I_n(\vec{r}_{xy})$, orders cannot simply be separated using the inverse mixing matrix $\bar{M}^{-1}$, as the sample information varies from image to image, i.e. the Fourier transformed sample is multiplied by different phase gradient, which after multiplication with $\bar{M}^{-1}$ may interfere to generate an amplitude modulation of the Fourier data.

5.4.1 Typical artefacts

If sample drift does occur between images, the components used to describe the formation of one image will change from image to image. Applying the inverse mixing matrix $\bar{M}^{-1}$ to the image vector therefore will not yield correctly separated components and the image reconstruction will fail.

While small drifts could in principle be tolerated, they may complicate the determination of other reconstruction parameters, in particular ones using cross-correlations (CC) between components for parameter optimisation (see 5.6).

5.4.2 Remedy: drift-correction through cross-correlation

Sample drift can be determined through cross-correlating the individual images $I_n(\vec{k}_{xy})$. This would normally be achieved by iteratively shifting one of the two images and trying to maximise the centre cross-correlation peak. However, in SIM this cannot be done quite as straightforwardly, as each image contains two more of less independent structures of information: the sample information, shifted by $\Delta\vec{r}_m$, and the grating illumination, which is also shifted from image to image, but by a different amount $\vec{s}_m$. Both these structures will show up as peaks in the cross-correlation image, and the presence of one peak may shift the maximum position of the other.

We try to avoid this distortion of the sample cross-correlation by low-pass filtering the Fourier images prior to cross-correlation in order to suppress the frequencies around $\vec{k}_\mathrm{g}$, which carry most of the illumination pattern's information. The filters used are either Gaussian or $\mathrm{sinc}(\pi\vec{k}\cdot\vec{k}_\mathrm{g}/|\vec{k}_\mathrm{g}|^2)$. The latter corresponds to a real space convolution with a rectangular function of the width of one grating period, in an attempt to smear out this grating.

Problems

While the drift correction works reasonably well for small sample drifts, it does not reliably correct large drifts, or drift for very noisy data. If drift correction is used in the reconstruction, one has to be aware that for drift corrected images the effective grating position will be different from the intended one, so that the mixing matrix $\bar{M}$ has to be calculated for the corrected grating positions $\vec{s}_m - \Delta\vec{r}_m$. For large drifts it can happen that the thus corrected mixing matrix may no longer be invertible, i.e. its column vector may no longer be linearly independent.

5.5 Unknown grating period

In order to shift the separated components to their correct positions in frequency space, precise knowledge of the grating period (or the grating's k-vector in frequency space, $\vec{k}_{\mathrm{g}}$) is required. This is a parameter given by the experimental setup; dimensions of the device used for projecting the grating (i.e. the pixel size of an SLM or the grating constant of a physical grating) are usually known with some accuracy, and the corresponding k-vector $\vec{k}_{\mathrm{g}}$ can be deduced from this value and the magnification of the microscope.

However, calculating $\vec{k}_{\mathrm{g}}$ this way may not always yield a result of sufficient precision, as already very small deviations from the true value may lead to significant artefacts in the reconstructed image.

5.5.1 Typical artefacts

A small error $\Delta\vec{k}_{\mathrm{g}}$ in the grating vector $\vec{k}_{\mathrm{g}}$ will lead to the separated components being shifted to slightly incorrect positions in Fourier space – with an error of $n\Delta\vec{k}_{\mathrm{g}}$ for the shifted n^{th} component. When this information is recombined with the information of neighbouring components, all frequency information from overlap regions will be present at least two times (or more if there is overlap between more than two components) in a distance of $\Delta\vec{k}_{\mathrm{g}}$. This leads to typical beating artefacts with a beating frequency $\Delta\vec{k}_{\mathrm{g}}$. For small errors (sub-pixel in Fourier space) this beating will be visible as a slow brightness change over position $\vec{r}_{xy}$ in the reconstructed real space image.

Figure 5.1 illustrates this effect. Here the reconstruction was done assuming a grating of 520 nm period rather than the 500 nm used in the simulation of the data (Fig. 5.1a). The aforementioned brightness change is clearly visible when comparing the image to the correctly reconstructed one shown in Fig. 5.1b.

Another typical artefact stemming from wrong grating periods is a splitting of structures. Two examples of that are indicated by red arrows (Fig. 5.1a). Note how the dark vertical lines are split and exhibit another, bright line in their centre, whereas in the correctly reconstructed image these lines are not split. This is due to a phase

mismatch between the recombined components.

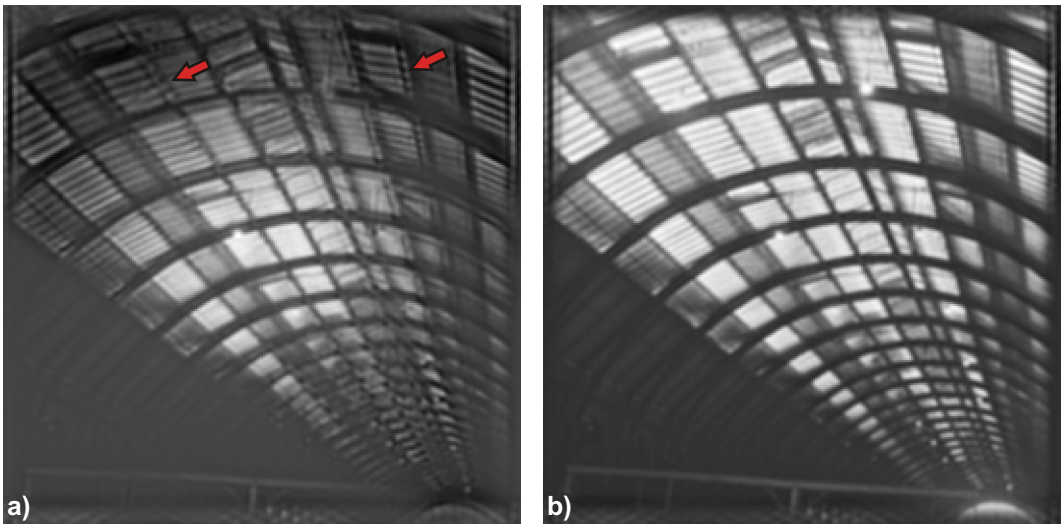


Fig. 5.1: Artefacts stemming from wrong grating period. (a) shows an image that has been reconstructed using the wrong grating period (all other parameters were assumed to be known). As a result, components are recombined under incorrect shifts, leading to beating artefacts. These can be seen as slow brightness changes over the image as compared to the correctly reconstructed image (b), i.e. image (a) is darker at its top and bottom edges. Another typical artefact stemming from wrong grating periods is a splitting of structures. Two examples of that are indicated by red arrows (a). Note how the dark vertical lines are split and exhibit another, bright line in their centre. This is due to a phase mismatch between the recombined components.

The error in grating period assumed in this example was relatively high (4%) in order to better illustrate typical artefacts. For smaller errors, this type of artefact may be barely noticeable. However, precise knowledge of the grating period is required at other steps of the reconstruction. There even small deviations from the accurate value may lead to severe artefacts.

Cross-correlations between different shifted components are used to optimise the mixing matrix $\bar{M}$ at a later stage of the algorithm (see 5.6). If the components are shifted incorrectly, this optimisation will not yield the correct results, leading to imperfectly separated components (Fig. 5.2a-c) and the corresponding artefacts (Fig. 5.3b).

Furthermore, recombination of components requires knowledge of the so-called global phase (the grating's absolute initial position). As this is also optimised through

cross-correlating shifted components, incorrect grating periods will lead to errors in the global phase and corresponding artefacts (Fig. 5.4).

5.5.2 Remedy: finding the grating vector through component cross-correlations

In order to find the exact grating k-vector $\vec{k}_\mathrm{g}$ we analyse the separated Fourier components $\vec{\tilde{\Omega}}^{(z_0)}$. The two shifted components $\tilde{\Omega}_i^{(z_0)}(\vec{k}+i\vec{k}_\mathrm{g})$ and $\tilde{\Omega}_j^{(z_0)}(\vec{k}+j\vec{k}_\mathrm{g})$ will in part contain the same information, if their respective shifted OTFs $\tilde{h}_i(\vec{k}+i\vec{k}_\mathrm{g})$ and $\tilde{h}_j(\vec{k}+j\vec{k}_\mathrm{g})$ overlap. If the shift vectors are chosen correctly this common information will be visible as a peak in the origin of the cross-correlation image between the two components. In order to optimise the shift vector, we iteratively vary $(i-j)\vec{k}_\mathrm{g}$ around the approximate expected value calculated from the experimental parameters (i.e. grating constant) and maximise the centre value of the weighted cross-correlation (WCC, see appendix A.8).

As can be seen from Fig. 4.3 it is not necessarily neighbouring components which have the most information overlap. On the other hand, some components may be attached to very weak illumination orders, resulting in cross-correlation values of low SNR despite high overlap. For the calculation of the grating vector components which yield cross-correlations with maximum SNR should be used.

Problems

At this point in the algorithm components may not yet be separated perfectly, i.e. if there are unknown parameters in the mixing matrix $\bar{M}$. We therefore cannot yet combine a stack of two-dimensional components $\tilde{\Omega}_n^{(z_0)}$ into one three-dimensional component $\tilde{\Omega}_n$, particularly as the individual slices may require different parameters for $\bar{M}$ or different global phase factors. Therefore we can also only correlate two-dimensional components, which makes the use of the ideal weights for a weighted cross-correlation impossible. Ideal weight-maps would have to be three-dimensional and would discard information from regions where the components' OTFs have no overlap.

Furthermore, as we need the exact grating k-vector for the iterative optimisation of $\bar{M}$ (see 5.6 to 5.7), we have to make do with components separated using an imperfect matrix $\bar{M}^{-1}$. The separated components may therefore not be perfectly separated but rather still contain residual information from other components. While these components are not good enough for image reconstruction, their cross-correlations nevertheless yield good results for the determination of the grating's k-vector $\vec{k}_{\mathrm{g}}$.

Gustafsson et al. [19] also use cross-correlations between components to find the exact k-vector. However, they do not allow for mixing matrix parameters to change between slices, which is justified by their very precise control of experimental parameters through active feedback mechanisms. This allows them to immediately separate and correlate three-dimensional components.

Other approaches of optimising the k-vector include the localisation of the Dirac spots – the spots visible at $\pm\vec{k}_{\mathrm{g}}$ in the Fourier transformed raw images $\tilde{I}_m(\vec{k}_{xy})$ – using multiscale products [64].

5.6 Fluctuations in grating phase

Component separation by means of the inverse mixing matrix $\bar{M}^{-1}$ can only be successful, if the matrix elements are known with high enough accuracy. Even small uncertainties in the grating position $\vec{s}_m$ in the individual images will lead to imperfect component separation and thus to artefacts in the final image.

Such uncertainties are particularly likely in systems in which the illumination grating is generated by a true physical grating and shifted by mechanically moving this grating, especially when this positioning is done without any active feedback. But even in systems where the grating is displayed and shifted by means of SLMs allowing precise knowledge of the true illumination phases, bleaching of the grating structure into the sample may lead to a shift in the *effective* illumination in the subsequent images.

5.6.1 Typical artefacts

Incorrect parameters in the mixing matrix $\bar{M}$ result in an imperfect separation of components. If components containing residual information of other components are shifted, this residual information will be shifted to incorrect frequencies in Fourier space. Most dominant is the information corresponding to the sample's zero-frequency. In the separated components this information will be present as peaks at multiples of the grating's k-vector $\vec{k}_{\mathrm{g}}$, as can be seen in Fig. 5.2a-c. Recombining these component leads to a Fourier image with these peaks still present, as shown in Fig. 5.3a. An inverse Fourier transform of just these six dominant peaks would yield a honeycomb pattern, and indeed the real space image shows such artefacts (Fig. 5.3b).

5.6.2 Remedy: optimisation of mixing matrix

In order to be able to correctly separate the individual components in light of fluctuations in grating phase the grating positions $\vec{s}_m$ in the mixing matrix $\bar{M}$ need to be estimated by the reconstruction algorithm. Our approach is to iteratively vary the matrix parameters and optimise for best separation of components.

Evaluation of separation quality using cross-correlations of components

Before they are shifted to their true frequency-space positions each correctly separated Fourier component should normally not contain any information from other correctly separated Fourier components. The central value of the weighted cross-correlations ($\circledast_w$) between components (see appendix A.8)

$$c_{ij} = [\tilde{\Omega}_i(\vec{k}) \circledast_w \tilde{\Omega}_j(\vec{k})](0)$$

should therefore be low for different components $i \neq j$ (except for random correlations in the components' information and noise). For $i = j$ the correlation value will obviously always be extremely high, as both information and noise will be correlated.

Components calculated from unmixing with a non-optimised matrix however will

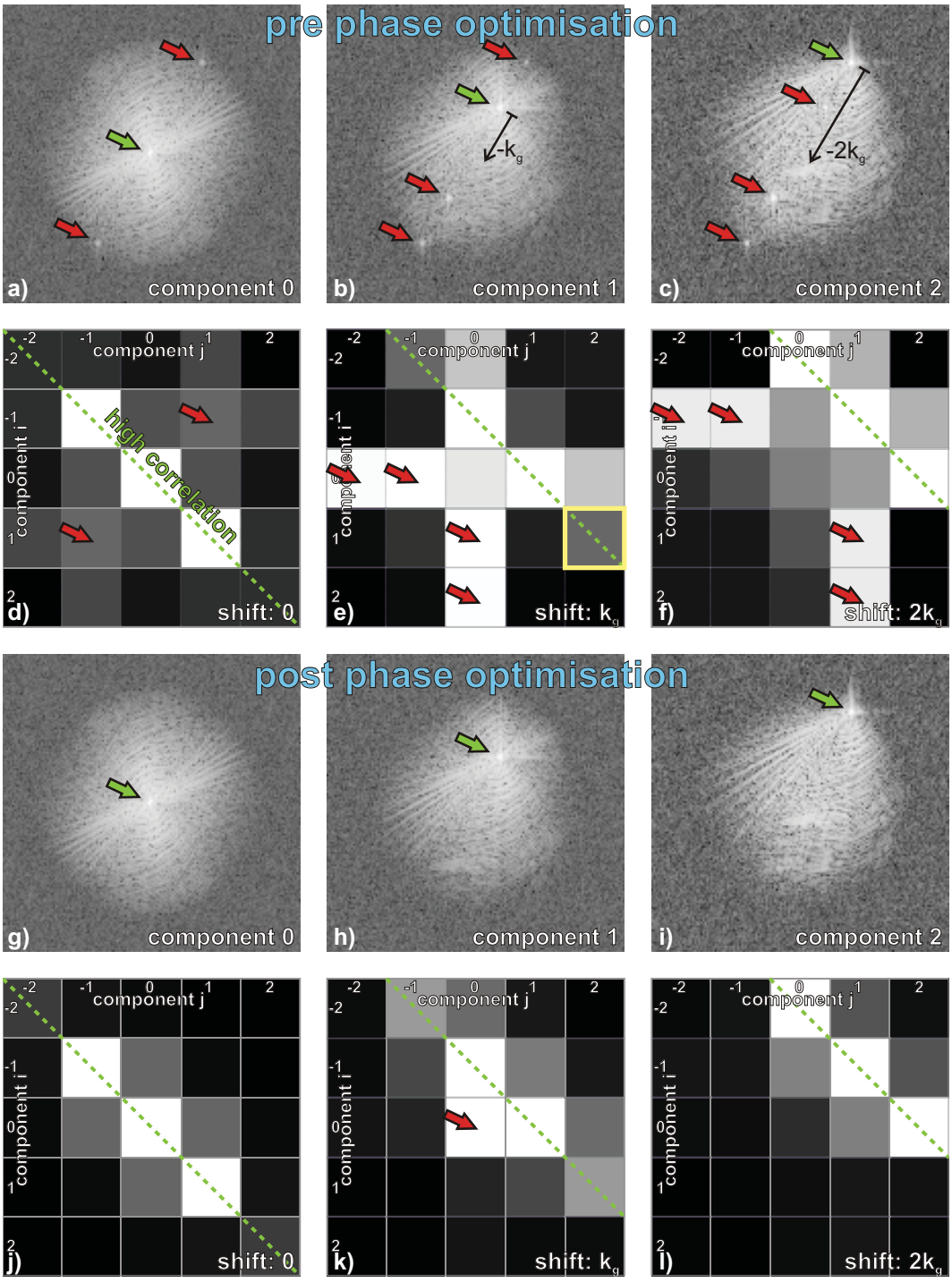


Fig. 5.2: Optimisation of grating phase. For a detailed explanation see 5.6.

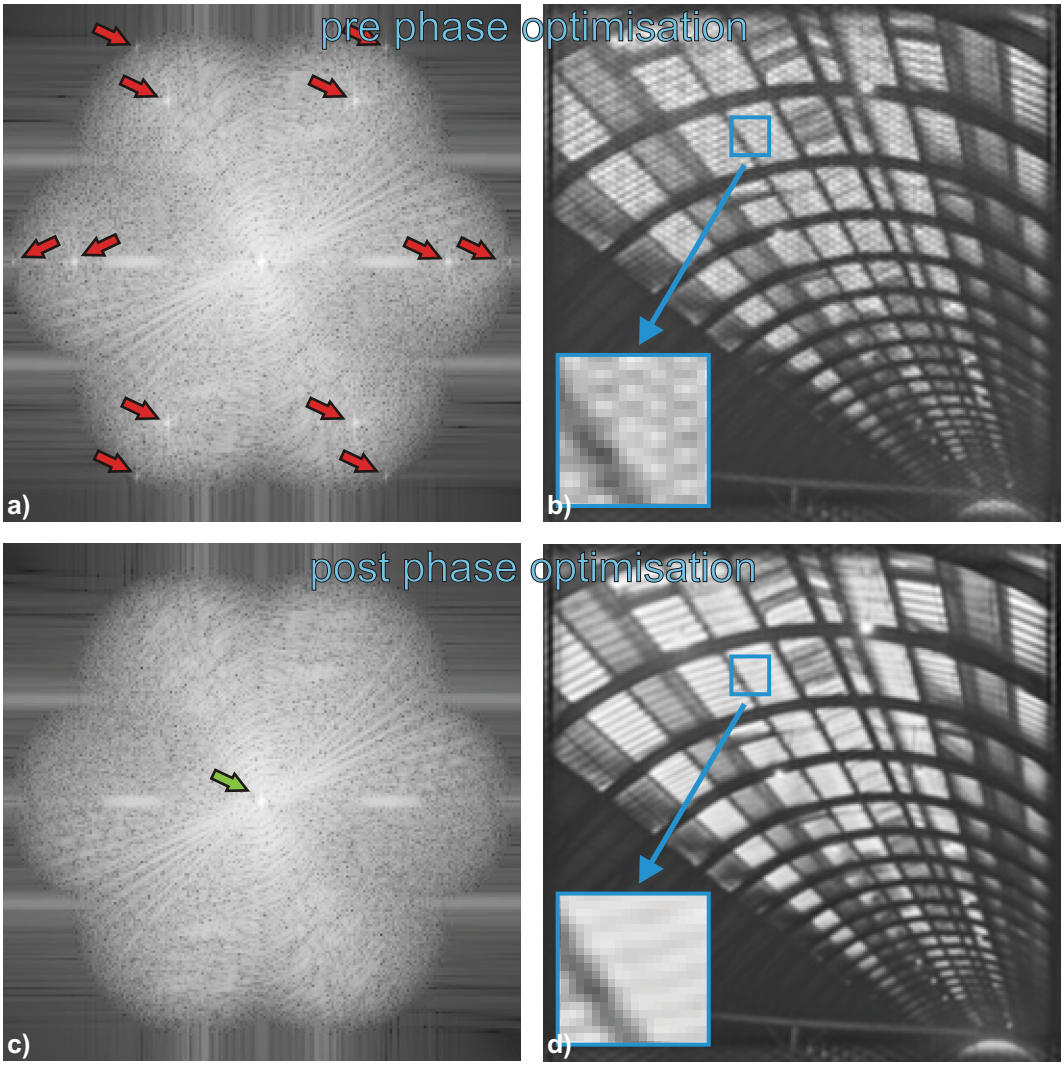


Fig. 5.3: As the result of imperfect separation residual sample information is translated to incorrect regions in Fourier space. This can be seen as peaks marked by red arrows (a) which correspond to the sample's zero frequency. As a result, the reconstructed real space image (b) exhibits periodic artefacts, which in the case of three SIM directions represent a honeycomb structure. Through an iterative optimisation of the mixing matrix $\bar{M}$ the residual information is removed from the separated components. The Fourier image (c) no longer shows incorrect peaks and the real space image (d) no longer exhibits periodic artefacts.

not be pure but rather still contain residual information from other components. In that case there will be overlap in information for different components $i \neq j$ and the cross-

correlation values c_{ij} will be higher than for the perfectly separated case[2]. Minimising the correlation values $c_{ij},\ i \neq j$ can therefore be used for optimising the mixing matrix $\bar{M}$.

Furthermore, when comparing separated components with components shifted by $-p$ grating k-vectors $\vec{k}_{\mathrm{g}}$, correlation values

$$c_{ij}^{(p)} = [\tilde{\Omega}_i(\vec{k}) \circledast_w \tilde{\Omega}_j(\vec{k} + p\vec{k}_{\mathrm{g}})](0) \tag{5.1}$$

should now be low for components $p \neq j - i$. For $p = j - i$ we expect high values, but unlike for the correlation between unshifted orders this is not because of the autocorrelation of noise; we therefore expect high values for good separations and can use this value in our optimisation.

This concept is illustrated in Fig. 5.2. Figures 5.2a-c show the components separated using an unmixing matrix $\bar{M}^{-1}$ without optimisation of phases. Figure 5.2a shows the separated zero order component, Figure 5.2b the first and Figure 5.2c the second order component. Residual components can clearly be seen in the peaks indicated by red arrows, whereas the desired peak in each component is marked by a green arrow. Figures 5.2d-f shows a graphic representation of the correlations $c_{ij}^{(p)}$ (Eq. 5.1) between the different components, where the row number indicate the component number $\tilde{\Omega}_i$, column number the components number $\tilde{\Omega}_j$ and Fig. 5.2d-f show correlation values for shifts of $-p\vec{k}_{\mathrm{g}},\ p = \{0, 1, 2\}$ of component $\tilde{\Omega}_j$. Grey levels indicate correlation strength, with white indicating strong correlations, black weak ones (percentile stretch display). Consider for example the square marked in yellow in Fig. 5.2e: here we have $i = 1$ ($\tilde{\Omega}_1$, Fig. 5.2b), $j = 2$ ($\tilde{\Omega}_2$, Fig. 5.2c) and a shift of $-\vec{k}_{\mathrm{g}}$, or $p = 1$. The grey level in this square therefore represents the correlation value $c_{1,2}^{(1)} = [\tilde{\Omega}_1(\vec{k}) \circledast_w \tilde{\Omega}_2(\vec{k} + 1\vec{k}_{\mathrm{g}})](0)$.

While for correlations between unshifted components ($p = 0$, Fig. 5.2d) we expect

[2] In some cases the correlation values for incorrectly separated components $c_{ij},\ i \neq j$ may actually become lower than for the correctly separated ones. This can easily be seen when looking at the cross-correlation between e.g. $\tilde{\Omega}_0 + \tilde{\Omega}_1$ with $\tilde{\Omega}_0 - \tilde{\Omega}_1$. Although both terms contain the same information, their correlation will be low because the individual contributions have opposing signs. However, this problem is avoided by other requirement of the optimisation process.

only the correlation values on the diagonal (marked by a green dashed line) to be high, we can see that also other correlations are significant (indicated by red arrows). For correlations between unshifted components and components shifted by one (two) grating vectors $\vec{k}_\mathrm{g}$ (Fig. 5.2e, f), we expect only the first (second) off-diagonal (again marked green dashed) to contain high correlation values. Here the unsuccessful unmixing is even more obvious (red arrows).

After iterative optimisation of the grating phases in the unmixing matrix $\bar{M}^{-1}$, the correlation values are significantly improved[3] (Fig. 5.2j-l) and residual components are no longer discernible in the separated components (Fig. 5.2g-i).

The residual peaks are also removed from the reconstructed Fourier image (Fig. 5.3c), yielding a final real space image (Fig. 5.3d) without honeycomb artefacts.

Merit function for optimisation

The above comparison of unshifted components with ones shifted by $p\vec{k}_\mathrm{g}$ can be done as long as the support of the two components still exhibits any overlap after shifting. This is the case for values of up to $p_0 = \max\{p:\ |p| < 2|\vec{k}_0|/|\vec{k}_\mathrm{g}|\}$. We therefore get a tensor C with dimensionality $(N, N, p_0 + 1)$ of correlation values $c_{ij}^{(p)}$ we can use for the optimisation of $\bar{M}$.

For our optimisation process we calculate a scalar figure of merit $F(C)$, which we try to minimise by iteratively varying the parameters of the mixing matrix $\bar{M}$. We empirically arrived at a merit function

$$F(C) = \sum_{i,j,p:p\neq j-i} |c_{ij}^{(p)}|,$$

which is simply a sum of all those correlation values' magnitudes, which are expected to be low. While this function may still be optimised further, we found it to give the most reliable results out of all functions tried.

[3] Figure. 5.2k still seems to show an unwanted high correlation (red arrow); however this correlation remained high even after reconstruction with perfect parameters and thus should also be high after phase optimisation.

Speeding up the optimisation

An optimisation following the above scheme would require iteratively changing the mixing matrix $\bar{M}$ and recalculating the component cross-correlation tensor C and the merit function $F(C)$. This means that $N^2p_0 - N(p_0+1) + (p_0^2 - p_0)/2$ cross-correlations values have to be calculated for each iterative step (the total number of elements in C less the number of elements on the on/off-diagonals which are supposed to be high), a feat that is very time-consuming especially for large images and in the case of nonlinear SIM with a large number of components.

This process can be significantly sped up. Rather than re-calculating the correlation tensor C after each iteration, one can calculate a different tensor D containing the correlation values of ($p\vec{k}_\mathrm{g}$-shifted) Fourier images, $d_{st}^{(p)} = [\tilde{J}_s(\vec{k}) \circledast \tilde{J}_t(\vec{k} + p\vec{k}_\mathrm{g})](0)$, containing M^2p_0 values. From this Fourier image correlation tensor D, the component correlation tensor C can be calculated via

$$C^{(p)} = \bar{M}^{-1} D^{(p)} \bar{M}^{-1\dagger}, \tag{5.2}$$

as is shown in appendix A.9. Here the dagger symbol † denotes the conjugate transpose or hermitian conjugate of a matrix. As D is independent of $\bar{M}$ it has to be calculated only once *before* the iterative optimisation of $\bar{M}$, requiring the one-time calculation of M^2p_0 cross-correlations, rather than the above mentioned $N(p_0+1) - (p_0^2 - p_0)/2$ per iteration. This dramatically speeds up the optimisation.

Performance

In the simulation the grating was moved with a nominal step size of 100 nm $\hat{=}\, 360°/5 = 72°$ between images (assuming a grating with a period of 500 nm in sample coordinates), 360° corresponding to a translation by a full period. Positional fluctuations were given as $\Delta\vec{s}_{1..5} = \{7.8\text{ nm}; -6.2\text{ nm}; 13.2\text{ nm}; -10.8\text{ nm}; -7.9\text{ nm}\} \,\hat{=}\, \{5.6°; -4.4°; 9.5°; -7.8°; -5.7°\}$. The phase fluctuations determined by the reconstruction algorithm were $\Delta\vec{s}'_{1..5} = \{0\text{ nm}; -14.2\text{ nm}; 5.4\text{ nm}; -19.0\text{ nm}; -15.8\text{ nm}\} \,\hat{=}\, \{0°; -10.2°;$

$3.9°; -13.7°; -11.4°\}$, leaving a difference of $\Delta\vec{s}_{1..5} - \Delta\vec{s}'_{1..5} = \{7.8\text{ nm}; 8.0\text{ nm}; 7.8\text{ nm}; 8.2\text{ nm}; 7.9\text{ nm}\} \mathrel{\hat{=}} \{5.6°; 5.8°; 5.6°; 5.9°; 5.7°\}$. After subtracting a global offset phase[4] of 7.9 nm $\mathrel{\hat{=}}$ 5.7°, this leaves a final phase error of $Err_{1..5} = \{-0.1\text{ nm}; 0.1\text{ nm}; -0.2\text{ nm}; 0.2\text{ nm}; 0.0\text{ nm}\} \mathrel{\hat{=}} \{-0.1°; 0.0°; -0.1°; 0.2°; 0.0°\}$.

While this impressively demonstrates the algorithms capability to correct for inaccuracies in the grating position, it also shows why these corrections are warranted: positioning inaccuracies of only around $\Delta\vec{s} \approx 10$ nm $\mathrel{\hat{=}}$ 7.2° (in sample space, physical displacements of the actual grating have to be multiplied with the magnification of the microscope objective) are enough to yield unsatisfactory reconstructions. The required positioning accuracies cannot easily be guaranteed without active feedback systems.

5.7 Fluctuation of order strengths between images

Like the overall intensity the relative strength of the individual orders may also change from image to image. This is most likely the case in nonlinear SIM [16, 25]. In the case of nonlinear SIM based on fluorescence saturation, a change in illumination intensity will directly translate into a change of individual order strengths. But also when using photo-switching the illumination dosage used for the switching process will affect the saturation of switching and thus the effective order strengths.

But even in linear SIM, where the true illumination orders have a fixed relative strength, the effective illumination may in fact exhibit slight variations in order strength. As an example consider linear SIM in the presence of photo bleaching: if the grating pattern is bleached into the sample structure, a combination of this bleached structure and the shifted patterned illumination may lead to a sample response that corresponds to an effective illumination containing higher orders than the true illumination. As these higher orders were not present the first image of the unbleached sample, their relative order strengths will clearly have changed between acquisitions.

[4] This phase optimisation can only correct for relative errors in grating position between images. After optimisation the components may still have a global phase factor, which is determined at a later stage (see 5.9).

5.7.1 Typical artefacts

As in 5.3 and 5.6 an imperfect separation of components leads to the corresponding artefacts in the final reconstruction, similar to those shown in Fig. 5.3b.

5.7.2 Remedy: optimisation of mixing matrix

Relative order strength variations between images can be incorporated into the mixing matrix $\bar{M}$ (Eq. 4.6):

$$\bar{M}_{mn} = v_{mn} \exp\{\imath n \vec{k}_{\mathrm{g}} \cdot \vec{s}_m\}.$$

These parameters v_{mn} can then be varied along with the grating phase parameters in the optimisation described in the previous section. As the zero order strength corresponds to the total intensity of the illumination and this intensity has been normalised to correct for fluctuations, the zero order strength is fixed at $v_{m0} = 1$.

Without this constraint, one would assume the optimisation routine to further optimise the image intensities, as the normalisation described in 5.3 does not account for possible fluctuations as a result of interaction between inhomogeneous illumination and sample. However, reconstructions of simulated data showed that – despite the excellent performance at optimising relative order strengths and phases – the iterative optimisation of illumination intensity does generally not yield better results than the initial normalisation (5.3).

Note that this optimisation does not yield any information about the absolute strengths of the individual orders, but rather only about relative changes thereof from image to image. The absolute order strengths are determined at a later stage (5.11).

5.8 Sample drift between different focal slices

Besides sample drift between the acquisition of individual images (see 5.4), sample drift can also occur between the acquisition of different focal slices.

5.8.1 Typical artefacts

If each slice can be individually optimised, the reconstruction of three-dimensional data acquired under sample drift (with changing focal position) will correspond to the imaging of a skewed sample with a skewed PSF.

5.8.2 Remedy: drift-correction through cross-correlation

While at this stage drift is more likely to be noticeable (slice acquisition happens on a slower time scale than the acquisition of individual SIM images), it is also easier to correct. Similar to the correction of sample drift between individual images (5.4), the different slices are aligned by iteratively shifting and cross-correlating the separated real space zero-order components (corresponding to the wide-field images) of the slices. Unlike the individual SIM images, these components should no longer exhibit grating patterns from the illumination, making them less prone to grating bias in the drift correction, thus voiding the necessity of Fourier-filtering components prior to drift-correction.

Problems

The iterative cross-correlation method used for correcting sample drift aligns images by maximising their overlap. In some rare cases this type of alignment may generate artificial sample drift rather than correct true sample drift. This can be the case if there are sample structures which are oriented at an oblique angle to the optical axis. Even without sample drift, different slice images may then exhibit similar image features, but with slight lateral translations. Although these translations are in fact correct, the drift correction will try to falsely align these images.

5.9 Unknown zero grating phase

The optimisation of the grating phases described in section 5.6 will only optimise the relative phase change between different images. If the initial phase or grating position

$\vec{s}_0$ is not known, the individual components will still be separated correctly, but may contain different global phase factors. This is true even in perfect systems, where the grating's absolute zero position also has to be determined.

5.9.1 Typical artefacts

If components with different global phase are combined using weighted averaging (section 4.2.5) their information content will start to interfere destructively. As a result the combined information content will be dampened (the magnitude of the average of two complex numbers of different phase is lower than the average of their magnitudes) and, worse, have the wrong phase. The dampening can be observed in Fig. 5.4e-h, where the

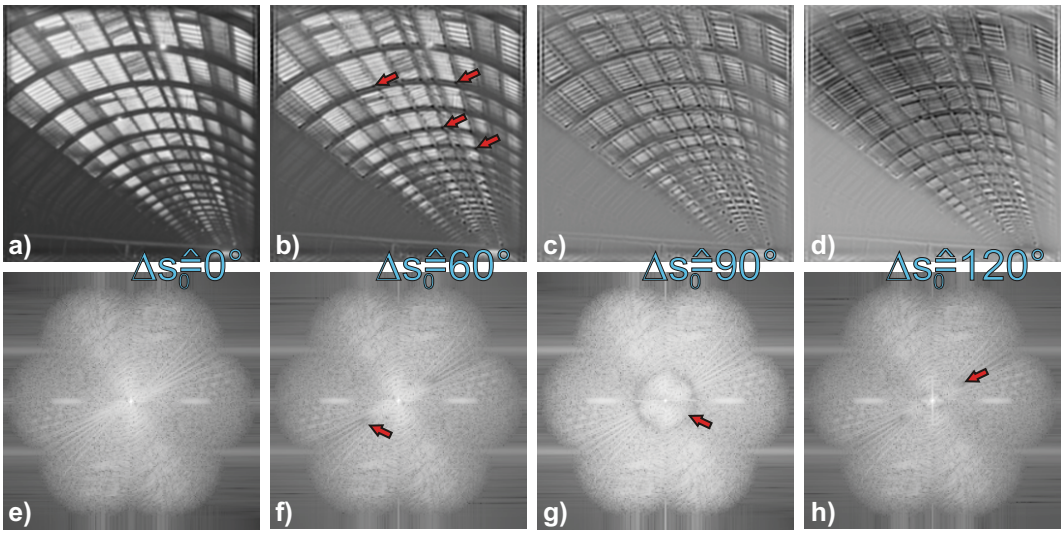


Fig. 5.4: Artefacts as a result of incorrect grating zero phase. Recombination of separated components with incorrect global phase will lead to interference between these components. In the recombined Fourier image this can be observed as a dampening of the information in the overlap regions. Examples of clearly visible dampening has been marked with red arrows f-h) for different global phase errors $\Delta\vec{s}$ corresponding to 60°, 90° and 120°. (b-d) show the corresponding real space reconstructions. Whereas the image corruption is obvious for larger phase errors (c,d), the artefacts are more subtle and thus more dangerous for smaller errors (b). For reference (a) and (e) show reconstructions using the correct global phase.

error in the grating's zero phase was 0°, 60°, 90° and 120° respectively. Regions where the dampening can be easily seen have been marked with red arrows. Figures 5.4a-d

show the corresponding real space reconstructions. Whereas the corruption of the images is quite obvious for the two largest errors (Fig. 5.4c, d), artefacts resulting from smaller errors (Fig. 5.4b) are more difficult to spot but therefore also more "dangerous". A few examples have been marked with red arrows.

5.9.2 Remedy: finding the global phase through cross-correlations

The global phase can simply be calculated from a weighted cross-correlation (see appendix A.8) between two shifted components. It is the negative phase of the cross-correlation value: $\Delta\vec{s}_{0,n} = \arg\{[\tilde{O}_n(\vec{k} + n\vec{k}_\mathrm{g}) \circledast_w \tilde{O}_0(\vec{k})](0)\}$. If the n^{th} component has no overlap with the zero-order component, its phase can be calculated either as a multiple of the first component's phase, $\Delta\vec{s}_{0,n} = n\Delta\vec{s}_{0,1}$, or through cross-correlating over a chain of components which do overlap, $\Delta\vec{s}_{0,n} = \arg\{[\tilde{O}_n \circledast_w \tilde{O}_{n-1}](0)\} + \Delta\vec{s}_{0,n-1}$.

5.10 Sample drift between rotational orientations

When acquiring three-dimensional image stacks the slowest mechanical operation is usually the rotation of the grating for acquiring the different directions. Therefore the order in which images are taken is usually as follows: first, M images are taken for M different grating positions under a single grating direction for one focus position z_0. Second, the sample is refocussed to a different focus position, and step one is repeated. Third, after a stack of images has thus been acquired, the grating is rotated to a new orientation, and steps one and two are repeated.

In the first setup built in our lab, grating rotation was so slow, that although sample drift could be neglected during the acquisition of a whole focal stack for one grating orientation, sample drift between the acquisition of different orientations could not.

5.10.1 Typical artefacts

As most parameters are optimised independently for each direction, sample drift between different directions will not affect this optimisation[5]. When combining all separated and shifted components for the reconstruction of the final image, sample drift between directions would show up in a distortion of the PSF and as motion-blurring in the final real space image. This blurring may void any resolution enhancement.

5.10.2 Remedy: drift-correction through cross-correlation

While at this stage drift is most likely to be noticeable (the acquisition of different directions happens on a much slower time scale than both the acquisition of different focal slices or individual SIM images), it is also easier to correct: after component separation, the separated zero-order components of all directions should ideally contain identical information. Any drift between them can be corrected similarly to the previously described drift correction (5.4), by iteratively shifting and cross-correlating the real space zero-order components. Unlike the individual images, the zero-order components do not exhibit a patterned illumination, making them less prone to errors in the correction, and also voiding the need to first Fourier-filter the components in order to avoid the grating bias.

5.11 Unknown order strengths

For the recombination of the individual components we have so far assumed that the effective, order-dependent OTFs $\tilde{h}_n$ (Eq. 4.5) are known, and with them the strengths of the illumination orders L_n (Eq. 4.1). And indeed for linear SIM these order strengths can be calculated reasonably well if the strengths of the amplitude orders entering the objective can be measured (e.g. for laser illumination).

However, there are situations when these "predicted" order strengths may not be

[5] For signal-to-noise reasons it can be beneficial to also compute cross-correlations of overlapping components from different directions for the optimisation of parameters. In that case, sample drift between directions may influence the parameter optimisation.

accurate enough. This is particularly the case for nonlinear SIM [16, 25], e.g. when the nonlinear behaviour of photo-switchable dyes (used for achieving the nonlinear sample response) is not precisely known, or when the strength of the nonlinear orders depends on the saturation of the fluorescence and therefore on the illumination intensity (and not just on the total dosage).

5.11.1 Typical artefacts

Incorrect order strengths do not result in significant artefacts per se. However, they will lead to an incorrect weighting of components in the weighted averaging recombination (4.2.5) and a thus recombined Fourier image will not have ideal signal-to-noise levels. Also, the magnitude of information in the Fourier image will be damped (or enhanced) if the strengths assumed for the weighted averaging are too high (too low). If high frequency are overemphasised this way, the reconstructed images may exhibit ringing artefacts, which can be reduced by digital filtering. Fortunately, phases are not affected.

Regarding image artefacts, this modification of information magnitude corresponds to a smooth modulation of the imaging OTF and thus also the PSF. This OTF modulation can magnify noise components resulting in corresponding artefacts.

5.11.2 Remedy: comparing different separated components

If precise values are not known, illumination order strengths can be estimated from the acquired data.

The basis for all estimations is that perfectly separated, shifted Fourier components share the same information where they overlap (except for noise contained in the Fourier components). This information is the Fourier transformed sample and is identical, except for a multiplication with the respective components' effective OTFs, $\tilde{h}(\vec{k}+n\vec{k}_\mathrm{g})$, which contain the order strengths. Therefore, we have

$$\tilde{O}_i(\vec{k}+i\vec{k}_\mathrm{g})\tilde{h}_j(\vec{k}+j\vec{k}_\mathrm{g}) = \gamma\tilde{O}_j(\vec{k}+j\vec{k}_\mathrm{g})\tilde{h}_i(\vec{k}+i\vec{k}_\mathrm{g}). \tag{5.3}$$

If the theoretical order strengths are correct, we get $\gamma = 1$. If not, $\gamma \neq 1$ gives us information about the true relative strengths of the orders.

There are several ways of calculating γ from the separated components. Our algorithm allows the choice of two different methods: first, a method based on comparing the total power contained in the frequency spectra of the separated components ("sum of squares"); and second, an iterative approach.

Sum of squares

For this method we compare the auto-correlations of both sides of Eq. 5.3. Because an auto-correlation does not cancel out noise as a cross-correlation between independent components would, the noise auto-correlation has to be subtracted from the component auto-correlation. We thus get

$$\gamma^2 = \frac{[(\tilde{O}_i\tilde{h}_j) \circledast (\tilde{O}_i\tilde{h}_j)](0) - \sigma_i^2[\tilde{h}_j \circledast \tilde{h}_j](0)}{[(\tilde{O}_j\tilde{h}_i) \circledast (\tilde{O}_j\tilde{h}_i)](0) - \sigma_j^2[\tilde{h}_i \circledast \tilde{h}_i](0)}.$$

Here functions' arguments were omitted for better legibility; the different functions are shifted in the same way as in Eq. 5.3.

Iterative approach

Starting from Eq. 5.3, the iterative approach tries to minimise the integral of absolute differences between both sides by varying γ:

$$\epsilon(\gamma) = \int (\Big||\tilde{O}_i\tilde{h}_j)| - |\gamma(\tilde{O}_j\tilde{h}_i)|\Big| \, d^3k.$$

This method was found to give the most accurate results.

Outlook

Both above mentioned techniques give biased results for noisy data. Formally the problem of finding the correct γ is a complex linear regression, where both datasets to be compared have independent noise. Although it has not yet been implemented in

our algorithm, the best solution is a technique used by Gustafsson et al. [17], which is described in appendix A.10.

6. IMAGES

The results of bead measurements and live cell imaging in this chapter have been published in [33] **L. M. Hirvonen, K. Wicker, O. Mandula and R. Heintzmann, *Structured illumination microscopy of a living cell*, European Biophysics Journal, 38 (2009), pp. 807–812**.

Our reconstruction algorithm has been used in various setups to achieve high resolution imaging. This chapter shows some examples of reconstructed SIM images.

6.1 Our setup

In our lab an SLM based system constructed by Liisa Hirvonen was used for various imaging tasks [31–33, 41]. A schematic diagram of this setup is shown in Fig. 6.1.

A continuous wave (CW) laser at 473 nm was used for fluorescence excitation. The laser on-time was controlled by a shutter (*Vincent-D1*, *Uniblitz*, New York, USA). The laser was focussed (lens L_1, focal length $f_1 = 25$ mm) through a 10 µm pinhole for mode cleaning, and then re-collimated (L_2, $f_2 = 50$ mm) to fill the screen of a phase-only spatial light modulator (*HEO 1080 P*, *Holoeye Photonics*, Germany), which was used to display grating patterns. The diffraction orders were focussed using lens L_3 and the focussed orders were then relayed (L_4, $f_4 = 150$ mm and L_5, $f_5 = 200$ mm) via a dichromatic mirror (*Comar*, Cambridge, UK) into the back focal plane of the microscope objective (63×/1.2, water immersion, *Leica*, Germany). Between the relay lenses, a $\lambda/2$-plate (*Comar*, Cambridge, UK) in a rotating stage (*DRTM 40*, *Owis*, Lübeck, Germany) and a polariser in another rotating stage ensured azimuthal polarisation of

the diffraction orders for maximum pattern contrast in the sample.

Light emitted by the sample was captured by the same objective, passed through the dichromatic mirror, was filtered by an emission filter and was then imaged via a tube lens (1.25×, *Leica*, Germany) onto a cooled CCD camera (*Imager intense*, *LaVision*, Germany).

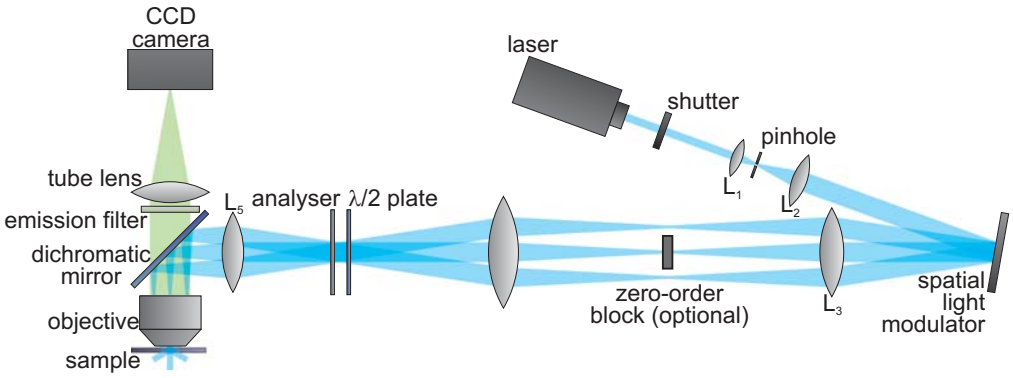


Fig. 6.1: Schematic diagram of the SIM setup developed by Liisa Hirvonen [33]. Note: This figure has been changed to correct an error in the original thesis submission.

6.1.1 Bead measurements

The system was characterised using measurements of green fluorescent polystyrene beads (*Duke Scientific*, California, USA; 71 nm specified bead diamter). For this 99 beads were extracted from an image, centred to sub-pixel accuracy and then averaged. These average images are shown in Fig. 6.2a for a wide-field and in Fig. 6.2b for the reconstructed SIM image. Figure 6.2c shows cross-sections through these average images (dashed) and Gaussian curves fitted to this data (solid). After accounting for the finite bead size of 71 nm diameter (see appendix A.11) the fitted curves yield PSFs with a full width at half maximum (FWHM) resolution of 249 nm for wide-field case and 96 nm for SIM.

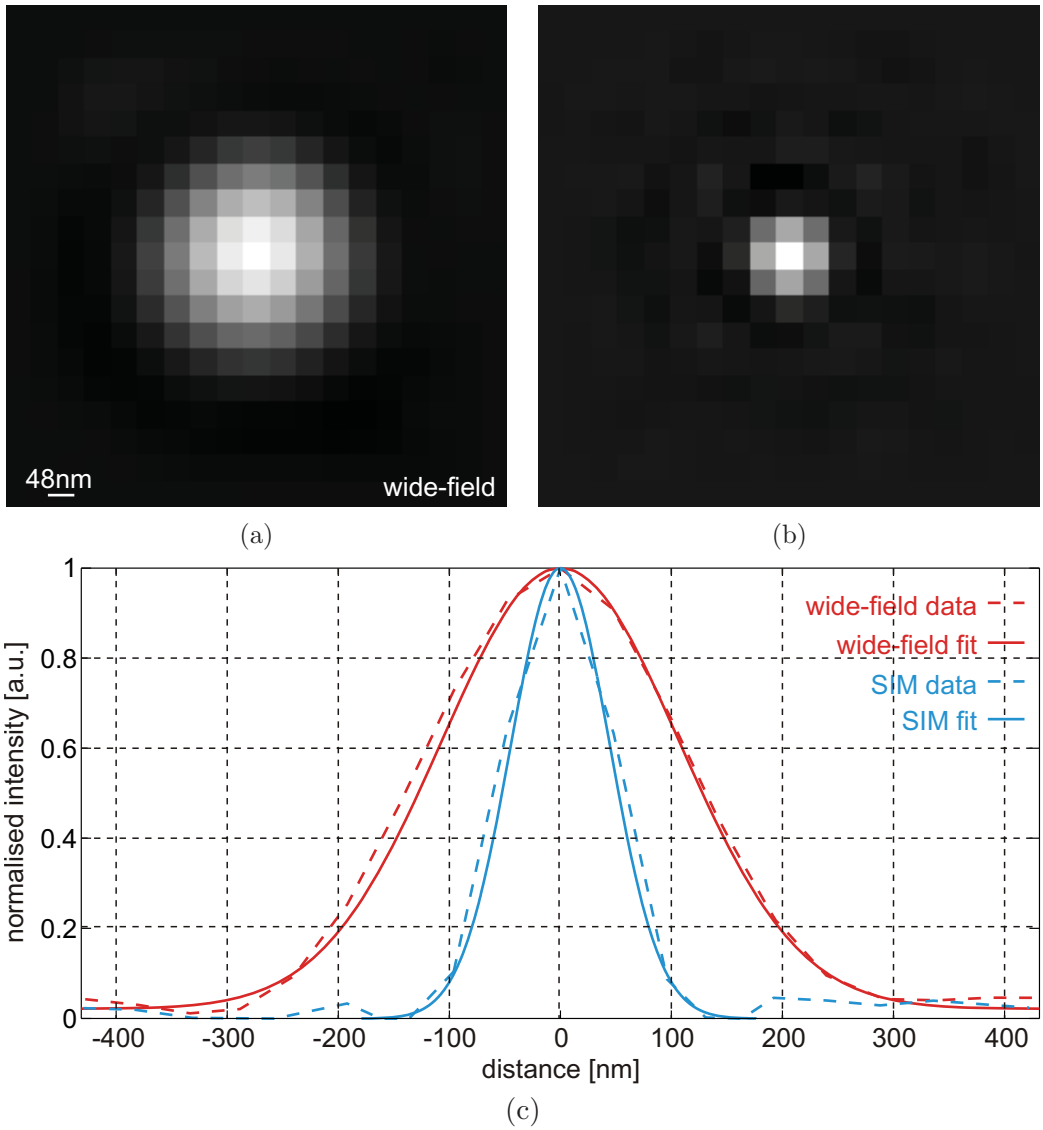


Fig. 6.2: Characterisation of the structured illumination microscope. (a) shows an average over 99 fluorescent beads of 71 nm diameter in a wide-field microscope, (b) in a reconstructed SIM image. (c) shows cross-sections through these bead images (dashed) and Gaussian curves fitted to this data (solid). After accounting for the finite bead size, the measured FWHM of 252 nm for wide-field and 105 nm for SIM yield an FWHM resolution of 249 nm for the wide-field PSF and 96 nm for the SIM PSF. Reproduced from [33].

6.1.2 Live cell imaging

Using this setup our group was one of the first to demonstrate structured illumination microscopy of a living cell. Figure 6.3a shows a single time-frame wide-field image of

MitoTracker-stained (*Invitrogen*, Paisley, UK) mitochondria in a living COS1 cell [14], Fig. 6.3b the corresponding reconstructed SIM image.

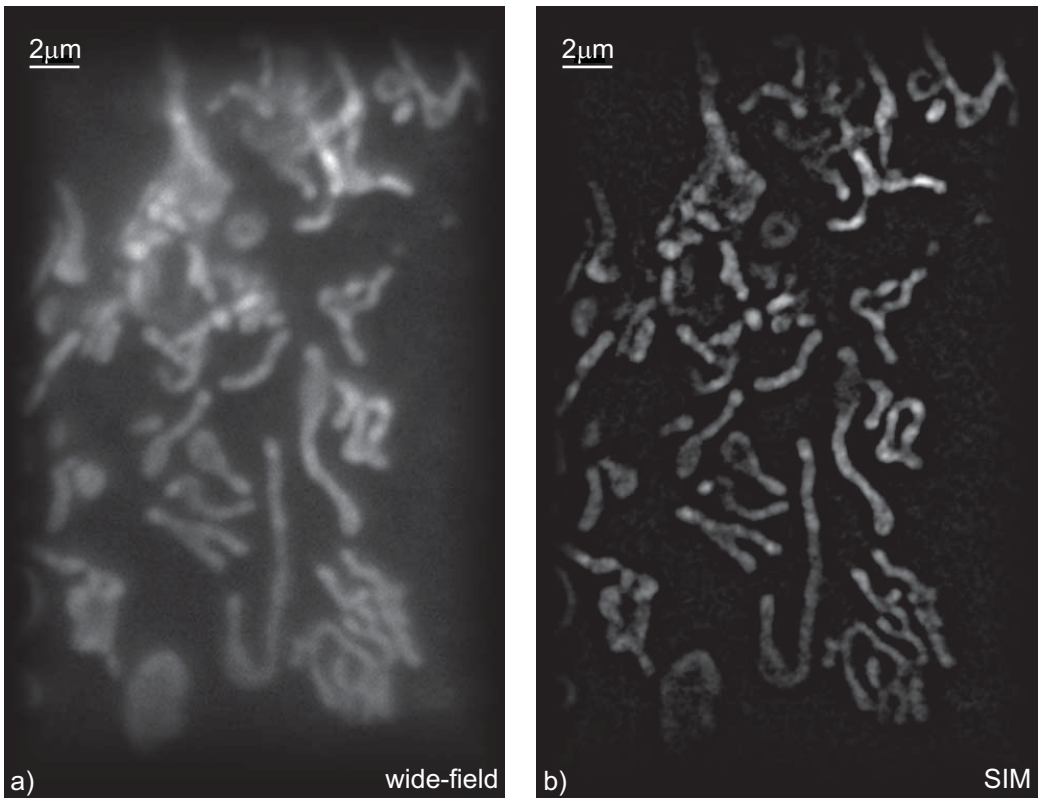


Fig. 6.3: *MitoTracker*-stained mitochondria in living COS1 cell. (a) shows the wide-field image, (b) the SIM image.

Time-frames were acquired at three minute intervals. Figure 6.4a shows an overlay of the first three wide-field frames, with the first, second and third frame being displayed in red, green and blue respectively. Figure 6.4b shows the same time series after SIM reconstruction. The SIM images show a clear improvement in resolution and sectioning over the wide-field images.

6.2 Zeiss SIM prototype

Our algorithms have been licensed by *Zeiss MicroImaging GmbH* (Jena, Germany) and are at the core of the image reconstruction of their new high-resolution microscope *Zeiss*

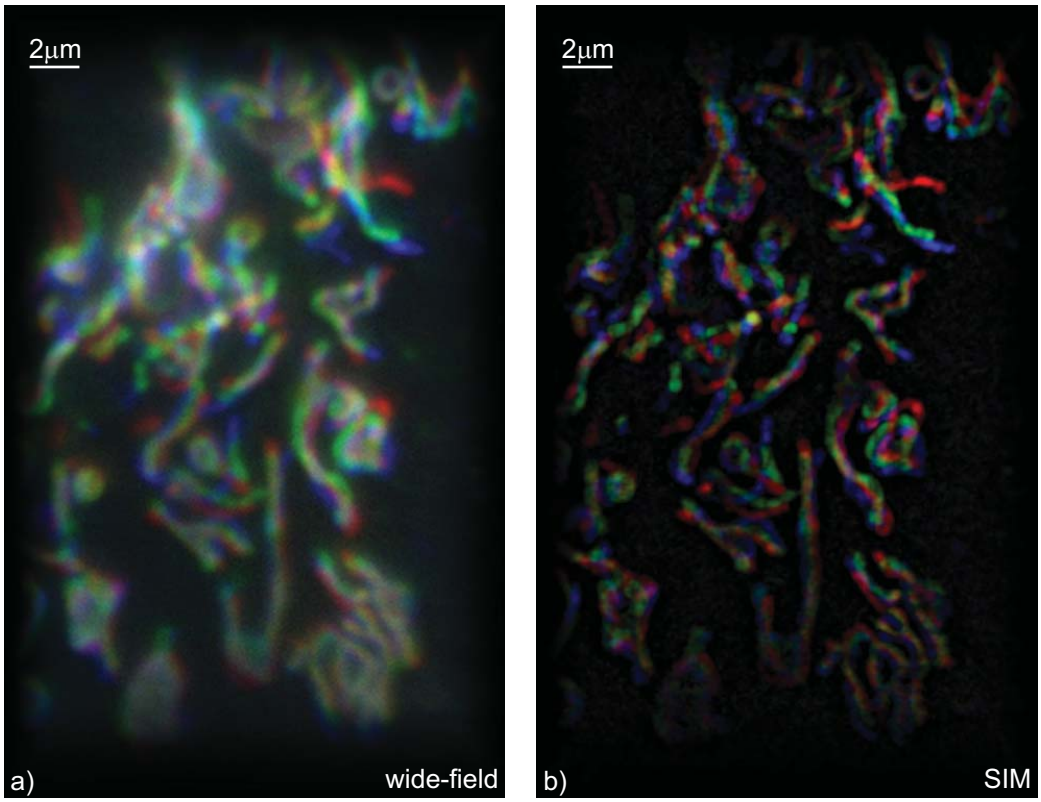


Fig. 6.4: Three frames of a time-series of *MitoTracker*-stained mitochondria in living COS1 cell. The indivual frames are displayed in red, green and blue respectively. (a) shows the wide-field series, (b) the SIM series.

Elyra S. Before commercialisation we were provided with a prototype of this system.

This prototype is based on the *Zeiss Axio Observer.Z1*, with a choice of four laser lines and several phase gratings. The measurements presented in this section were carried out using the following experimental parameters: 63×, 1.4 oil objective; illumination at 561 nm; phase grating with 29 lines/mm, corresponding to 73% of the maximum available k-vector; three grating directions; three-order illumination; five grating phases per acquired grating direction. Images were acquired using a cooled CCD camera (*Andor*, Belfast, UK).

6.2.1 NIH32 cells

Figure 6.5 shows a sample image acquired with this *Zeiss* prototype microscope. The sample used was an actin ruffle sample of NIH32 cells stained with *Phalloidin-RRX* and embedded in *ProLong Gold* (*Invitrogen*, Paisley, UK) which were prepared by Magdalene Michael of the *Randall Division of Cell and Molecular Biophysics*, *King's College London*. The images were acquired by Marie Walde [65].

The resolution improvement achieved by SIM is obvious when comparing Fig. 6.5b (wide-field) and Fig. 6.5c (SIM), which are close-ups of Fig. 6.5a (wide-field) and Fig. 6.5d (SIM). The improved resolution of SIM yields an image with finer detail as well as better contrast and sectioning.

Figure 6.6 shows a plot along the lines marked in light blue in Fig. 6.5b, c. Here the resolution enhancement becomes obvious: whereas the wide-field image shows only one peak (red), the SIM image clearly resolves two structures which are spaced approximately 250 nm apart.

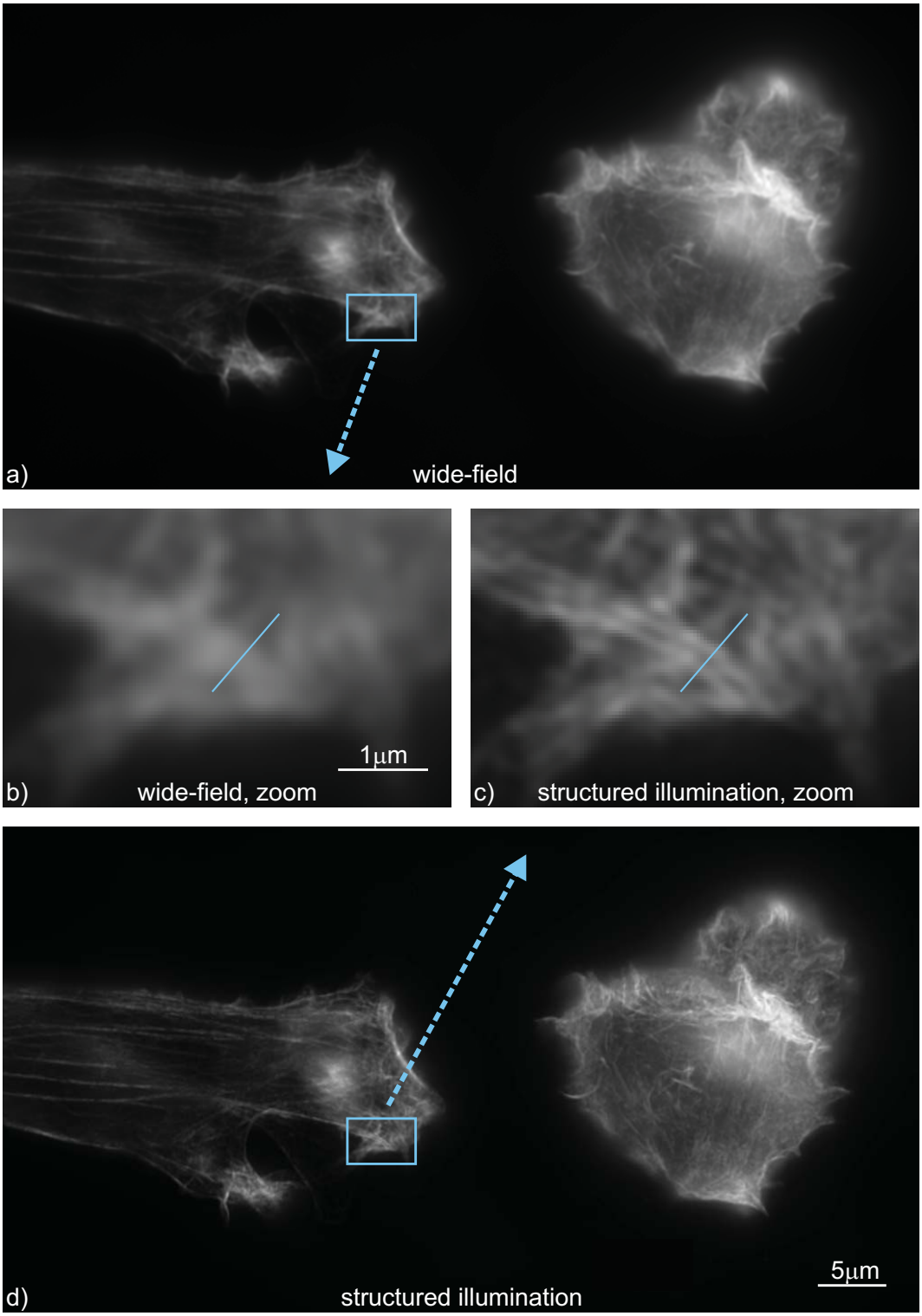


Fig. 6.5: Actin ruffle sample of NIH32 cells. (a) shows the wide-field image with a close-up (b), (d) the SIM image with a close-up (c). Data acquired by Marie Walde [65].

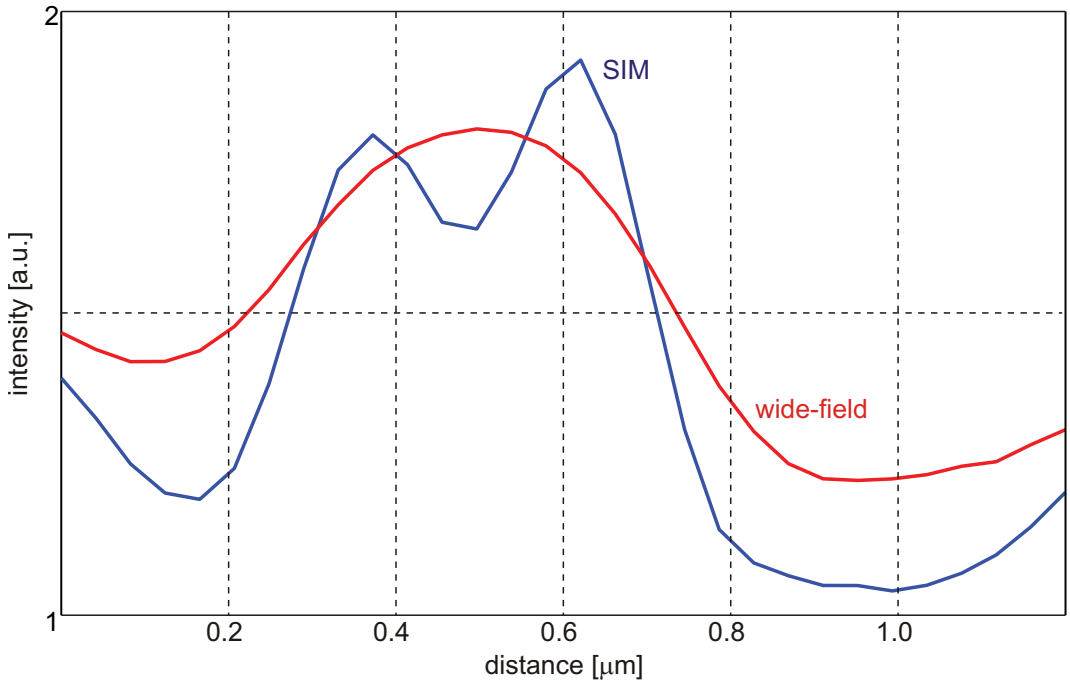


Fig. 6.6: Line plots along the light blue line in Fig. 6.5b (red, wide-field) and Fig. 6.5c (blue, SIM). Figure produced from data acquired by Marie Walde [65].

7. STRUCTURED ILLUMINATION MICROSCOPY: OUTLOOK

With microscopes such as the *DeltaVision OMX* by *Applied Precision* (Issaquah, Washington), the *Elyra S* by *Zeiss* or the *N-SIM* by *Nikon* (Kingston, UK) SIM is becoming a popular technique for the acquisition of high resolution microscopy images. Current commercial systems are still too slow for satisfactory live cell imaging. However, employing SLM based systems free of any moving parts SIM has been demonstrated to be capable of live cell imaging with frame rates of up to 11 Hz [38]. The resolution enhancement of linear SIM is limited to a factor of about two, but recent advances in nonlinear SIM [16, 31] suggest that SIM may well become a viable alternative to super-resolution techniques such as STED, PALM, STORM, etc.

While for sophisticated systems guaranteeing precise knowledge of all experimental parameters reconstruction of SIM images from the acquired data is somewhat straightforward [19], it is challenging to construct systems of such quality. The reconstruction algorithms developed in our group facilitate SIM imaging on less sophisticated, home-built systems [33] and are at the core of the commercial *Zeiss Elyra S*.

For both SNR and resolution it may be beneficial to reconstruct SIM data using maximum likelihood deconvolution [21] techniques adapted for this purpose and we are working to extend our algorithms to allow this type of reconstruction as well as reconstruction in real space rather than Fourier space, which may be less prone to artefact generation.

We are confident our current and future algorithms will enable researchers to reliably reconstruct images acquired on their own systems and help make SIM more accessible to the scientific community.

Part III

RESOLUTION AND EFFICIENCY ENHANCEMENT USING IMAGE INVERSION INTERFEROMETERS

8. INTRODUCTION

Despite recent developments in the field of super-resolution [5, 16, 19, 30, 54] confocal microscopy [2] remains the work-horse of the biological community. It is a robust technique, applicable to a wide variety of samples without too stringent requirements on sample preparation or selection of dyes. Through the combination of laser-scanning with a pinhole, out-of-focus light is blocked, resulting in optically sectioned images of the sample. Besides sectioning, the pinhole can also lead to an improvement in resolution compared to wide-field microscopy. In the case of a (nearly) closed pinhole, the final PSF is the product of two conventional wide-field PSFs (at different wavelengths), nearly doubling the frequency support. However, a smaller pinhole also leads to a loss in signal, as not only out-of-focus light is blocked. The result is a reduction in signal, which can render any resolution enhancement effectively useless.

A related technique, multi-photon excitation microscopy [7], relies on the nonlinearity of multi-photon excitation to achieve optical sectioning without a pinhole. The longer wavelengths used are also advantageous when imaging in deep tissue, as there is less chance of unwanted absorption or scattering, a fact that is being exploited in clinical diagnostics [39]. At the same time, however, the longer wavelengths reduce the microscope's resolution capability.

As will be shown in this part of this thesis, image inversion interferometry can be used to enhance the lateral resolution and/or detection efficiency of the above methods.

8.1 Confocal microscopy

Unlike wide-field microscopy or related methods such as structured illumination microscopy, confocal microscopy is a scanning technique. For each scan position the

sample is illuminated with a focused spot, the illumination PSF h_{illu} of the system. To calculate the PSF of this system, let us consider a fixed scan position and a point source at a position $\vec{r}$ in sample space. The illumination intensity at this position is $h_{\mathrm{illu}}(\vec{r})$. The emitted light is then imaged into the pinhole plane of the optical system, which is described by a convolution of the point source with the detection PSF h_{det}. A pinhole is placed in this image plane of the system, and whatever light lies inside the boundary of the pinhole (expressed by a multiplication with the pinhole function p), is usually collected by an integrating detector, such as a photomultiplier tube (PMT) or an avalanche photodiode (APD). Mathematically the above can be expressed as (see appendix B.1)

$$\begin{aligned} h_{\mathrm{total}}(\vec{r}) &= h_{\mathrm{illu}}(\vec{r})[h'_{\mathrm{det}} \otimes p](\vec{r}), \\ \tilde{h}_{\mathrm{total}}(\vec{k}) &= \tilde{h}_{\mathrm{illu}} \otimes [\tilde{h}'_{\mathrm{det}}\tilde{p}](\vec{k}) \end{aligned} \tag{8.1}$$

where $h'_{\mathrm{det}}(\vec{r}') := h_{\mathrm{det}}(-\vec{r}')$ is the mirrored detection PSF. If we assume a round pinhole of diameter d, $p(\vec{r}') = \Pi(r'_x/d, r'_y/d, \delta(r_z))$ (where Π is the unit rectangular function), Eq. 8.1 yields

$$\tilde{h}_{\mathrm{total}}(\vec{k}) = d^2\tilde{h}_{\mathrm{illu}}(\vec{k}) \otimes [\tilde{h}'_{\mathrm{det}}(\vec{k})\frac{J_1(d\sqrt{k_x^2 + k_y^2})}{\sqrt{k_x^2 + k_y^2}}],$$

where J_1 is the first order Bessel function of the first kind. This equation shows us, that while a small pinhole diameter d^2 leads to an enhancement of higher lateral frequencies $k_x,\ k_y$ in the Bessel term, the overall detection efficiency decreases, an effect of the multiplication with d. In fact, as d approaches zero it becomes the dominant term in the above equation, and no light will be detected. Regarding only the in-focus part of the PSF ($z = 0$), this effect is illustrated in Fig. 8.1 for various pinhole sizes. In Fig. 8.1a the normalised OTFs clearly show that smaller pinholes yield a higher relative transmission efficiency of high frequencies than larger pinholes. However, as the unnormalised OTFs in Fig. 8.1b make clear, the absolute transmission efficiency decreases for smaller pinholes.

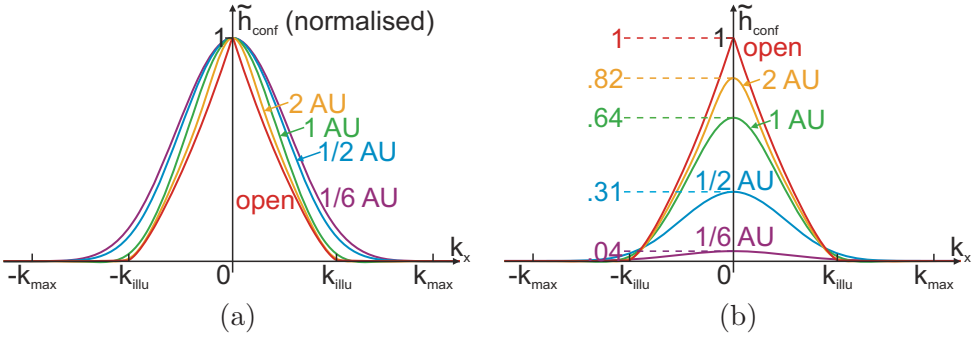


Fig. 8.1: Confocal in-focus OTFs for various size pinholes. Smaller pinholes yield a better relative transmission of higher spatial frequencies, as can be seen when the OTFs are normalised to the same overall transmission (a). However, smaller pinholes lead to more light being discarded and the overall transmission decreases (b). Although the relative transmission of high frequencies may be higher, the absolute transmission may actually be decreased, as can be seen by comparing the OTFs for 1/2 and 1/6 Airy Unit pinholes in the normalised and the unnormalised picture. k_{illu} shows the maximum spatial frequency available in the illumination, whereas k_{max} is the support of the combined illumination and detection.

8.2 Multi-photon excitation microscopy

Unlike confocal microscopy, two- and multi-photon excitation scanning microscopes [7] do not require a pinhole to achieve optical sectioning, but use the nonlinear dependency between illumination intensity and absorption probability. For an N-photon process, this absorption probability scales with the illumination intensity to the power of N. We therefore get an excitation PSF $h_{\mathrm{ex},N}(\vec{r}) = h_{\mathrm{illu}}^{N}(\vec{r}/N)$, where the scaling of the position $\vec{r}$ with $1/N$ is due to the fact, that to excite a fluorophore with a given excitation wavelength λ_{ex} via a simultaneous absorption of N photons, these photons need to be of wavelength $\lambda_N = N\lambda_{\mathrm{ex}}$, and that the PSF scales with the wavelength. The emitted light is then usually collected using an integrating detector without a pinhole, so that $h_{\mathrm{ex},N}$ is also the PSF of the whole N-photon imaging process.

In Fourier space this yields an OTF of $\tilde{h}_{\mathrm{ex},N}(\vec{k}) = [\bigotimes_{n=1}^{N} \tilde{h}_{\mathrm{illu}}](N\vec{k})$, where $\bigotimes$ denotes multiple convolutions. The support of this N-photon OTF is independent of N: while the N-fold auto-convolution of $\tilde{h}_{\mathrm{illu}}$ has an N times larger support than $\tilde{h}_{\mathrm{illu}}$ alone, this is countered by the scaling of the frequency vector, $N\vec{k}$. However, while

they have the same support, the resolution nevertheless worsens for greater N, as the higher frequencies get weaker with every auto-convolution.

8.3 Enhancing resolution and efficiency

Although confocal microscopy nearly doubles the frequency support compared to conventional wide-field microscopy, the improvement of lateral resolution is not very strong. Significant enhancement of the lateral resolution can only be achieved for very small pinholes, in which case the detection efficiency drops rapidly, resulting in noisy images [1].

Multi-photon excitation microscopy does not yield higher resolution, but it also suffers from low efficiency, as the probability of multi-photon excitation is very low.

As will be shown in the next chapter, using image inversion interferometers, lateral resolution can be significantly improved for same-sized or even larger pinholes (or in the case of multi-photon excitation microscopy, without pinhole). As a result, the detection efficiency of confocal microscopes can be enhanced, as the same resolution may be obtained for significantly larger pinholes.

[1] Note that the main purpose of a confocal microscope, namely optical sectioning, can be achieved for relatively large pinholes without much loss in signal.

9. RESOLUTION AND EFFICIENCY ENHANCEMENT USING IMAGE INVERSION INTERFEROMETERS

Parts of this chapter including figures have been published in [67] **K. Wicker and R. Heintzmann, *Interferometric resolution improvement for confocal microscopes*, Optics Express, 15 (2007), pp. 12206–12216**.

9.1 The idea

In 2006 Sandeau et al. proposed the 4Pi' microscope [56, 57] as a way to increase the lateral resolution of regular 4Pi microscopes [28]. It is a modification of the 4Pi microscope that contains an image inversion system in one of the microscope's arms. But although essential in the 4Pi' microscope, the separate arms of the 4Pi microscope are not necessary for this technique, as the same principle can be applied to regular confocal microscopes.

To do so, an interferometer is placed in the descanned detection pathway of the microscope, as shown for the Mach-Zehnder interferometer in Fig. 9.1. The light distribution coming from the microscope is split at the first beam splitter. It is then laterally inverted in one of the interferometer's arms before being recombined at the second beam splitter. After passing through optional pinholes, both the interferometer's constructive (I_+) and destructive (I_-) output intensities are measured.

Figure 9.2 illustrates the working principle of image inversion interferometry. If light is emitted by a source on the optical axis (Fig. 9.2a, b), the inverted field distribution will be identical to the non-inverted one and the two can interfere. Therefore all

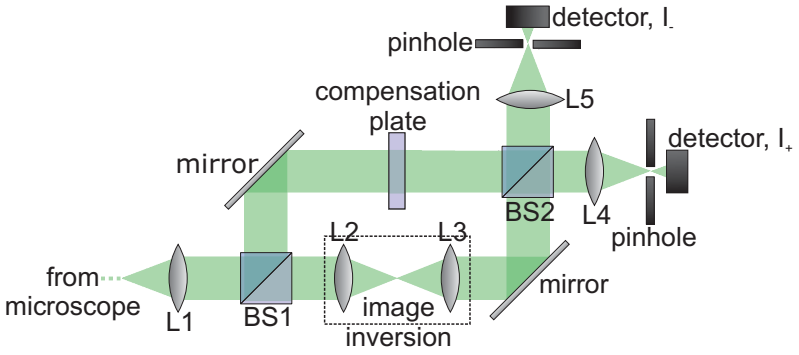


Fig. 9.1: Image inversion interferometer. Light coming from a microscope is split, inverted in one arm of the interferometer and recombined.

light will be collected in the constructive output (Fig. 9.2a), while the destructive output remains dark (Fig. 9.2b). If the source is slightly off-axis (Fig. 9.2c, d), the two field distributions will still have significant overlap and will interfere, albeit not perfectly. This leads to a decreased signal in the constructive output (Fig. 9.2c) and an increased signal in the destructive output (Fig. 9.2d). For sources far off-axis (Fig. 9.2e, f) the two field distributions will have hardly any spatial overlap and therefore will not interfere, leading to equal signals in the constructive (Fig. 9.2e) and destructive (Fig. 9.2f) channel. This general bias of on-axis light being detected preferably in the constructive channel leads to an improvement in lateral resolution over conventional detection.

9.2 Theory

9.2.1 The in-focus point spread function

In order to calculate the detection PSF (full-field illumination, no pinhole) of the image inversion interferometer, we assume a two-dimensional (lateral only[1]) system and analyse the integrated signal emitted by a point source positioned at $\vec{r}$. In image space this point source will generate an amplitude distribution $a(\vec{r}'_{xy} - \vec{r}_{xy})$, where $a(\vec{r}'_{xy})$ is the coherent or amplitude PSF in image coordinates. Interference with its inverted copy

[1] As will be shown later, this detection PSF is independent of axial position z.

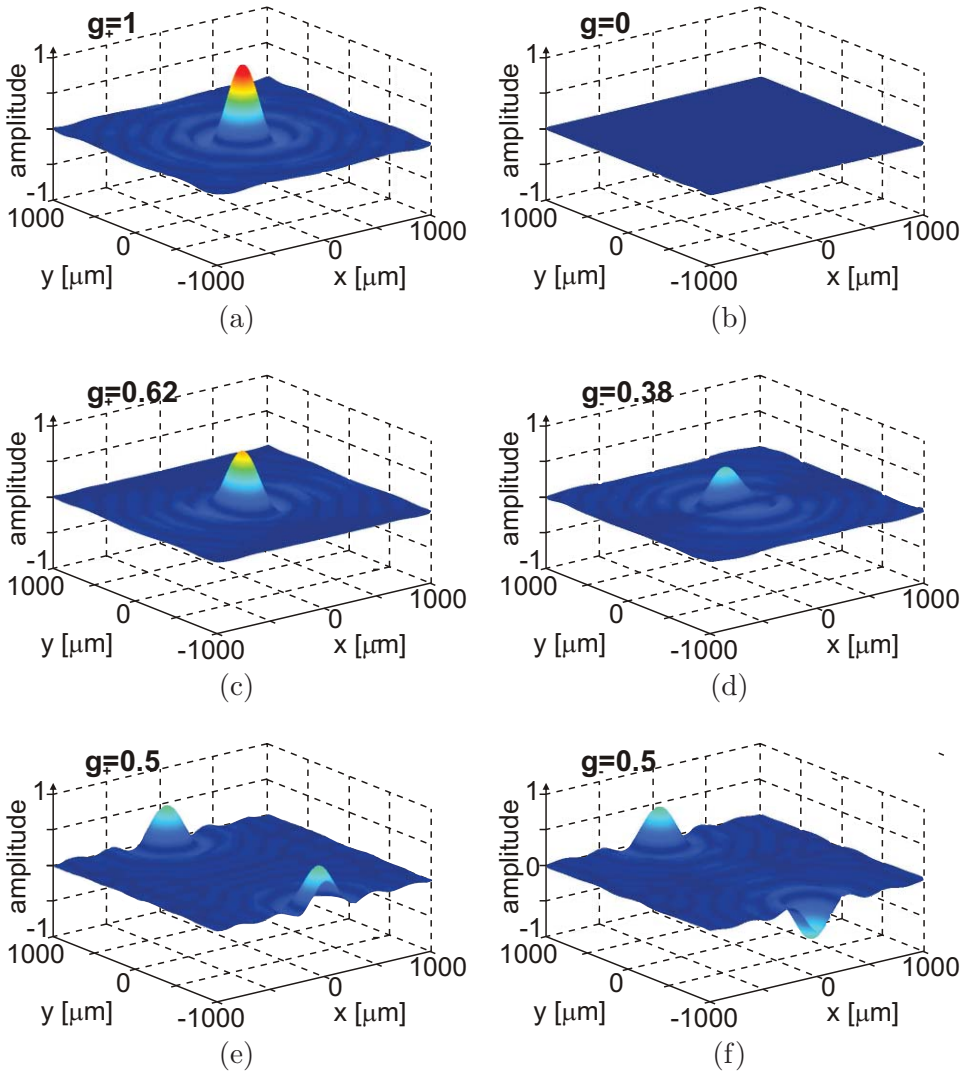


Fig. 9.2: Interference of coherent point spread functions in the image inversion interferometer. (a, c, e) show the recombined field distributions in the constructive interferometer channel, (b, d, f) in the destructive channel. For a source on the optical axis (a, b) the non-inverted and inverted field distributions are identical and can therefore interfere. After integration this leads to a full signal $g_+ = 1$ in the constructive channel (a) and zero signal $g_- = 0$ in the destructive channel (b). If the point source is close to the optical axis the two field distributions will have significant overlap and still interfere, albeit not perfectly. For the (arbitrary) distance of 160 nm shown in (c,d), this leads to a signal of $g_+ = 0.62$ in the constructive channel (c) and $g_- = 0.38$ in the destructive channel (d). For large distances from the optical axis (e, f) the two distributions have hardly any spatial overlap and therefore cannot interfere. This leads to equal signals of $g_+ = g_- = 0.5$ in both channels.

and integration over the whole image plane (open pinhole) yields (see appendix B.2)

$$g_{\pm}(\vec{r}_{xy}) = 1/2(1 \pm [a \otimes a^*](-2\vec{r}_{xy})).$$

The index $\pm$ denotes the constructive $(+)$ and destructive $(-)$ channel, respectively. For the final detection PSF we will be subtracting the destructive signal from the constructive one:

$$g_{\mathrm{det}}(\vec{r}_{xy}) = g_+(\vec{r}_{xy}) - g_-(\vec{r}_{xy}) = [a \otimes a^*](-2\vec{r}_{xy}). \tag{9.1}$$

This eliminates the constant offset in $g_{\pm}$. The corresponding OTFs are

$$\tilde{g}_{\pm}(\vec{k}_{xy}) = 2\pi^2\delta(\vec{k}_{xy}) \pm (1/8)\tilde{a}(-\vec{k}_{xy}/2)\tilde{a}^*(\vec{k}_{xy}/2) \tag{9.2}$$

for the constructive and destructive channels, where the δ-peak is responsible for the constant offset in the signal, and

$$\tilde{g}_{\mathrm{det}}(\vec{k}_{xy}) = (1/4)\tilde{a}(-\vec{k}_{xy}/2)\tilde{a}^*(\vec{k}_{xy}/2) \tag{9.3}$$

for the difference signal. For the case of symmetrical systems this interferometric OTF is the absolute square of the original ATF magnified to twice its original frequency range, as indicated by the argument $\vec{k}_{xy}/2$. This is an improvement even over the non-interferometric wide-field OTF, $\tilde{h}_{\mathrm{wf}}(\vec{k}_{xy}) = \tilde{a}(\vec{k}_{xy}) \otimes \tilde{a}^*(-\vec{k}_{xy})$, which is an auto-correlation of the original ATF and drops to zero at the edges of the OTF's support. However it is worth mentioning that for full-field illumination and without a pinhole (which we assumed for the calculation of the interferometric PSF/OTF) the confocal detection PSF is constant and therefore has no resolution at all, whereas the wide-field resolution can only be achieved for a closed pinhole, in which case no light is collected.

As an example we look at the conventional wide-field ATF in scalar theory at low

numerical aperture:

$$\tilde{a}(\vec{k}_{xy}) = \begin{cases} \frac{1}{\sqrt{\pi}k_c} & , |\vec{k}_{xy}| \leq k_c \\ 0 & , |\vec{k}_{xy}| > k_c \end{cases}$$

This yields an interferometric OTF

$$\tilde{g}_{\pm}(\vec{k}_{xy}) = \begin{cases} 2\pi^2\delta(\vec{k}_{xy}) \pm \frac{1}{8\pi k_c^2} & , |\vec{k}_{xy}| \leq 2k_c \\ 0 & , |\vec{k}_{xy}| > 2k_c \end{cases} \tag{9.4}$$

and a corresponding interferometric PSF

$$g_{\pm}(\vec{r}_{xy}) = \frac{1}{2}\left(1 \pm \frac{J_1(2k_c|\vec{r}_{xy}|)}{2k_c|\vec{r}_{xy}|}\right) \tag{9.5}$$

where J_1 is the first order Bessel function of the first kind. The difference signal is

$$g_{\mathrm{det}}(\vec{r}_{xy}) = \frac{J_1(2k_c|\vec{r}_{xy}|)}{(2k_c|\vec{r}_{xy}|)}. \tag{9.6}$$

Figure 9.3a shows the interferometric detection OTFs $\tilde{g}_{\pm}$ and $\tilde{g}_{\mathrm{det}}$ in comparison to the corresponding wide-field OTF $\tilde{g}_{\mathrm{wf}}$. The improvement is obvious: while $\tilde{g}_{\pm}$ and $\tilde{g}_{\mathrm{det}}$ have the same support as $\tilde{g}_{\mathrm{wf}}$ they do not fall off towards the edge of the support region, therefore significantly enhancing the higher frequency components. This effect on the resolution can be immediately seen in Fig. 9.3, which shows the corresponding PSFs. Although no pinhole was assumed, the resolution of the difference signal is far better than that of the conventional wide-field signal, which can only be achieved for a closed pinhole.

9.2.2 Out-of-focus behaviour

To calculate the out-of-focus behaviour of the interferometric detection PSF/OTF, we can defocus the ATF by multiplying it with the z-dependent free-space ATF $\tilde{o}_z(\vec{k}_{xy}) =$

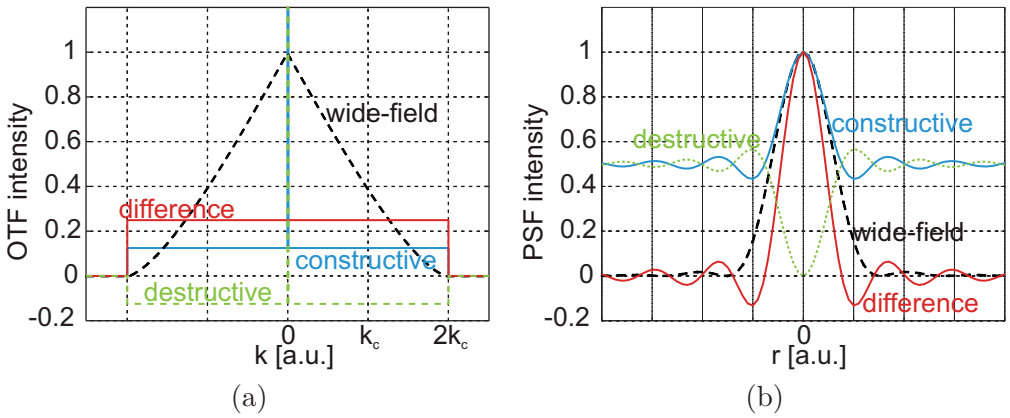


Fig. 9.3: Comparison of OTFs (a) and PSFs (b). The resolution improvement is strongest for the difference signal $g_{\mathrm{det}} = g_+ - g_-$. The interferometric OTFs do not fall off towards the edge of the support region, therefore enhancing high frequency components. Note that the interferometric signals were calculated for detection without pinhole, in which case the detection PSF (OTF) of a confocal system would be constant (a δ-peak) and not contribute to the overall resolution at all.

$\exp\{-\imath z\sqrt{(2\pi/\lambda)^2 - |\vec{k}_{xy}|^2}\}$ [55]. Equation 9.2 then becomes

$$\begin{aligned}\tilde{g}_{\pm,z}(\vec{k}_{xy}) &= 2\pi^2\delta(\vec{k}_{xy}) \pm (1/8)\tilde{a}(-\vec{k}_{xy}/2)\tilde{o}_z(-\vec{k}_{xy})\tilde{a}^*(\vec{k}_{xy}/2)\tilde{o}_z^*(\vec{k}_{xy}) \\ &= 2\pi^2\delta(\vec{k}_{xy}) \pm (1/8)\tilde{a}(-\vec{k}_{xy}/2)\tilde{a}^*(\vec{k}_{xy}/2)|\tilde{o}_z(\vec{k}_{xy})|^2 \\ = \tilde{g}_{\pm}(\vec{k}_{xy}),\end{aligned}$$

and therefore remains unchanged, because $\tilde{o}_z(\vec{k}_{xy})$ is phase-only with an amplitude of one. The interferometric detection-only PSF/OTF is thus independent of the axial position z.

9.3 Minimising the constant background

While it is remarkable that image inversion interferometry achieves a relatively high resolution without a loss of on-axis signal, the 50% offset in the signal is undesirable in imaging. There are, however, several ways to reduce or eliminate this constant background in the detection PSF.

9.3.1 Subtraction of signals

As the PSFs of the two interferometer outputs differ only in the sign of the non-constant terms, subtracting the two signals yields the difference signal $g_{\mathrm{det}} = g_+ - g_-$ from which the offset has been removed, as previously mentioned and shown in Fig. 9.3. However, as the two original signals are detected independently, their noise will also be independent and cannot be removed by the subtraction of the signals. So while off-axis sources will no longer contribute to the signal, they will continue to contribute to noise. Also, one has to consider that the difference signal g_{det} may contain negative values. This can be avoided by doing a scaled subtraction, or by using more sophisticated approaches like weighted averaging in Fourier space [26].

9.3.2 Localised illumination

The contribution of the constant background (i.e. the δ-peak in the OTF $g_\pm$) can also be reduced by combining interferometric detection with localised illumination (e.g. confocal [2] or two-photon excitation [7]). The final PSF is the product of illumination and detection PSF. As the illumination PSF drops to zero away from its centre, so will the product of the two PSFs, removing the constant background.

Translated into Fourier space the argument is that as long as the illumination OTF does not have a δ-peak the combined OTF, which is the convolution of the detection OTF with the illumination OTF, will no longer exhibit a δ-peak, either. Interferometrically enhancing high frequencies in the detection OTF results in an improved frequency response in the combined OTF.

Furthermore, even without a pinhole the combination of interferometric detection and scanned illumination leads to optical sectioning, as the convolution of the detection OTF with the wide-field OTF of the illumination fills the missing cone.

9.3.3 Detection pinholes

Introducing detection pinholes also reduces unwanted contributions caused by the constant offset in the detection PSF, as the pinhole will block light originating from points

far off the optical axis, the same way it does in confocal microscopes. Additionally, the pinholes also block out-of-focus light and therefore improves confocal sectioning. At the same time the relative significance of interferometric detection on the lateral resolution will be reduced for smaller pinholes: for the case of a closed pinhole the PSF g_+ resulting from a symmetric APSF a is identical to the regular confocal PSF, while g_- remains zero and no interferometric improvement in resolution is achieved. The influence of pinhole size on the performance will be discussed in section 9.4.3.

9.3.4 Multiple use of the interferometer

The constant offset in Eq. 9.5 can in principle also be eliminated through multiple use of interferometers. Light in the constructive output can be subjected to another pass through the same or a similar interferometer. However, applying the inversion operation twice results in the identity operation $\hat{1}$. Again using image inversion in a second interferometer would therefore always lead to a signal in the constructive output only. A different operation has thus to be used in order to reduce the offset. When using a second interferometer, an image rotation of 90° would reduce the offset to 1/4. As long as the operation $\hat{R}$ applied on the image fulfils the conditions $\hat{R}^n \neq \hat{R}^m$ and $\hat{R}^n a(0) = a(0),\ \forall n, m \in \mathbb{N},\ n \neq m$, multiple use of the interferometer can make the constant offset arbitrarily small while preserving the on-axis performance.

9.4 Simulations

Using *MATLAB* (*The MathWorks*, Massachusetts, USA) together with the *DIPimage* toolbox (Quantitative Imaging Group, TU Delft, The Netherlands) we simulated the effect of interferometric detection in combination with confocal illumination and detection pinholes. The simulations were done for circularly polarised light using high NA vector theory [50]. Interference was calculated individually for each electric field vector component. Parameters used for these simulations were: excitation wavelength $\lambda_{\mathrm{ex}} = 488$ nm, detection wavelength $\lambda_{\mathrm{det}} = 525$ nm, numerical aperture $NA = 1.2$, refractive index $n = 1.33$. In order to distinguish the resulting PSFs from the detection-

only PSFs we refer to the simulated PSFs as h_C (confocal), $h_\pm$ (interferometric with pinhole) and $h_{dif} = h_+ - h_-$ (difference in interferometric outputs).

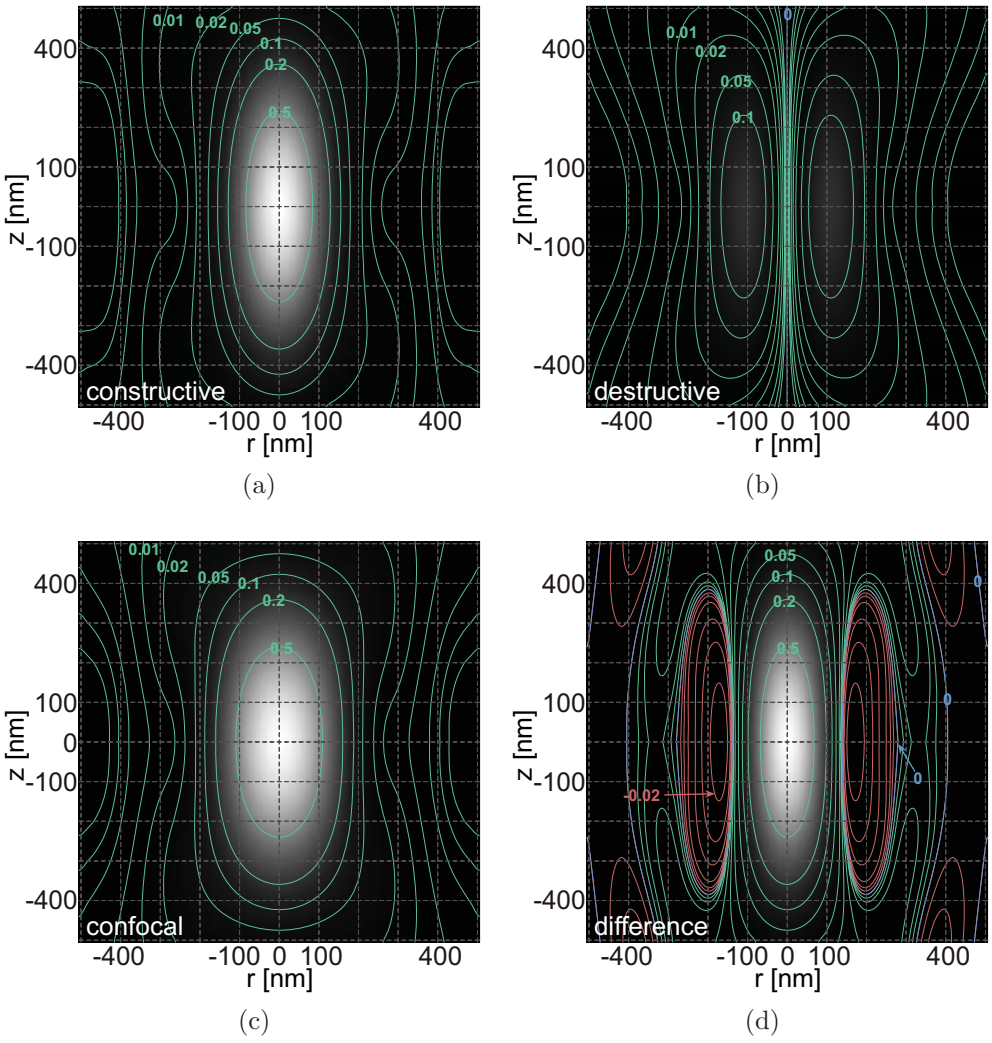


Fig. 9.4: PSFs for the constructive (a) and destructive (b) interferometer outputs $h_\pm$, for the confocal case without an interferometer h_C (c) and the difference in signal in the two interferometric outputs h_{dif} (d). Interferometric detection yields an improvement in lateral resolution: Lateral FWHM are 218 nm for h_C, 168 nm for h_+ and 135 nm for h_{dif}. h_{dif} exhibits small negative values of about -2.9% of the PSF's maximum. For simulation parameters see section 9.4. In all cases the pinhole has the size of one Airy disc.

9.4.1 Point spread functions

Figure 9.4 shows the resulting PSFs for a pinhole of the size of one Airy disc. Figure 9.4a shows the constructive output h_+. While the axial resolution along the optical axis is the same as for the confocal case h_C (Fig. 9.4c), the lateral resolution improvement is evident, with FWHM of about 168 nm for h_+ as compared to 218 nm for h_C. Subtracting the destructive output signal h_- (Fig. 9.4b) from h_+ results in the difference signal h_dif shown in Fig. 9.4d. This eliminates the δ-peak in the OTF Eq. 9.4 and therefore the offset of 1/2 in Eq. 9.5, thus further improving the resolution, to a FWHM of about 135 nm. Note that this subtraction results in some negative values (red lines) of -2.9% of the PSF's maximum.

9.4.2 Sectioning capability

The axial FWHM of the interferometric PSF (with detection pinhole and confocal illumination) is governed by the pinhole size and not improved beyond that of the corresponding confocal PSF. However, the sectioning capability is still improved: already for a completely open pinhole the interferometric method achieves sectioning, as the final OTF $\tilde{h}_\mathrm{dif}$ is a convolution of the detection OTF $\tilde{g}_\mathrm{dif}$ and the wide-field illumination OTF $\tilde{h}_\mathrm{wf}$ and thus fills the missing cone.

Figure 9.5 shows the simulated signal generated by a homogeneous fluorescent plane perpendicular to the optical axis for a conventional confocal microscope $C_\mathrm{C,plane}$, as recorded in the constructive interferometric output $C_\mathrm{+,plane}$ and for the difference in constructive and destructive signal $C_\mathrm{dif,plane}$. While for large distances from the focus all curves exhibit the same falloff proportional to z^{-2}, the interferometric curves $C_\mathrm{+,plane}$ and $C_\mathrm{dif,plane}$ drop off faster for small distances from the focus, leading to a FWHM of about 617 nm for $C_\mathrm{C,plane}$, 594 nm for $C_\mathrm{+,plane}$ and 550 nm for $C_\mathrm{dif,plane}$. At 1 µm from the focal plane the suppression in the difference signal is improved by a factor of about three over the confocal signal.

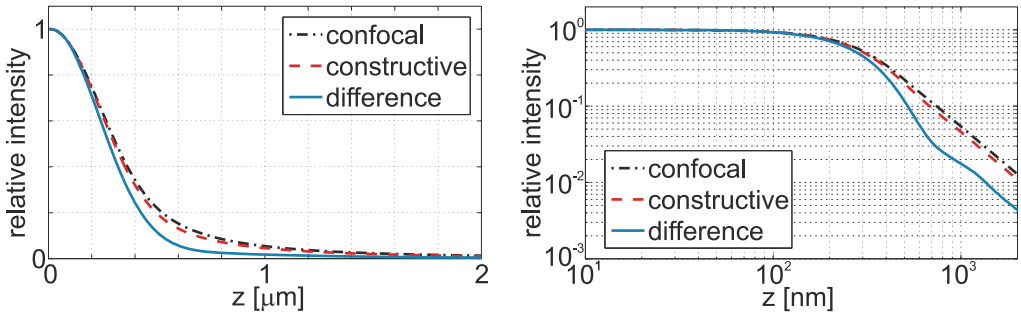


Fig. 9.5: Capability of the various methods for sectioning fluorescent planes. (a) The interferometer has an increased sectioning capability, with the difference signal surpassing the constructive output. (b) The logarithmic plot shows a z^{-2}-dependence far away from the focal plane for all methods. However, the intensities of the interferometric measurements fall off more quickly close to the focus.

9.4.3 Influence of pinhole size on resolution and sensitivity

Lateral resolution

While the use of a pinhole is needed in order to achieve good optical sectioning, it also reduces the resolution enhancing effect of the interferometric detection. This enhancement is strongest for a completely open pinhole and vanishes as the pinhole is closed completely. Figure 9.6a shows the dependence of the lateral FWHM resolution in the focal plane on the size of the pinhole. The simulation parameters used were the same as for the previous PSF simulation (9.4). For the confocal microscope (red) the resolution steadily worsens with increasing pinhole size until approximately 1.4 Airy Units, from whereon the resolution is dominated by the illumination and stays nearly constant. The signal detected in the constructive interferometer channel (turquoise) has a resolution which is nearly independent of the pinhole size and corresponds to the resolution of a confocal microscope with a closed pinhole. For the difference signal (blue) the resolution actually improves with increasing pinhole size. Its best resolution of around 130 nm (open pinhole) is about 24% better than the highest possible resolution of a confocal (170 nm for a closed pinhole) and around 43% better than a confocal with a similarly open pinhole (230 nm, same as wide-field).

Signal-to-noise ratio

The resolution of a confocal microscope can thus be significantly enhanced. Furthermore, this enhancement is achieved for larger pinholes which means that more light will be collected and the signal levels can potentially be improved. For a fixed pinhole size the peak signal from an in-focus point source is the same for the confocal, constructive and difference case. The constructive channel however rejects more of the unwanted light not stemming from the momentary scan position. This results in a better SNR for the constructive output than for the confocal case. This can be seen in Fig. 9.6b, where the peak signal of an in-focus point source over the noise stemming from a 3D fluorescent sea [44] has been plotted against pinhole size. The simulation parameters used were the same as for the previous PSF simulation (9.4). As the variance of the signals is always additive, the SNR of the difference channel will be the same as for the confocal case, which corresponds to a sum of constructive and destructive signals.

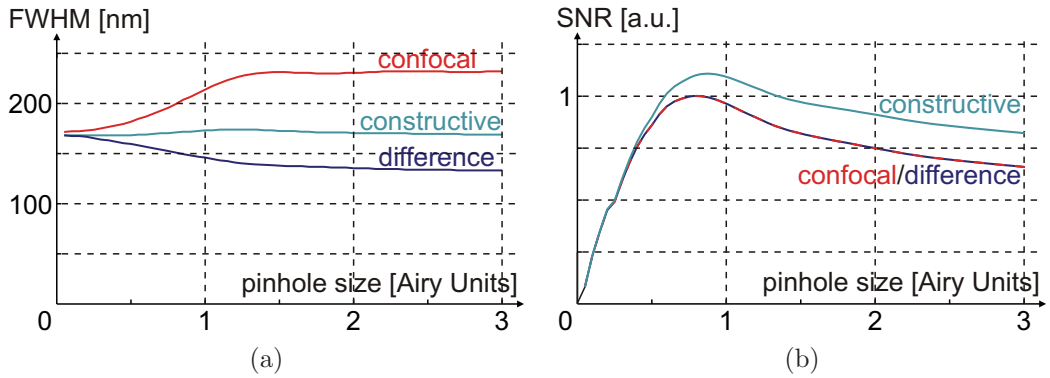


Fig. 9.6: (a) Lateral FWHM resolution plotted against pinhole size. (b) Signal-to-noise ratio plotted against pinhole size; peak signal of a point source over shot-noise from a 3D fluorescent sea. The simulation parameters used were the same as for the previous PSF simulation (9.4).

10. POINT SPREAD FUNCTION MEASUREMENTS

Parts of this chapter including figures have been published in [70] **K. Wicker, S. Sindbert and R. Heintzmann, *Characterisation of a resolution enhancing image inversion interferometer*, Optics Express, 17 (2007), pp. 15491–15501**.

The setup we had originally suggested in order to illustrate the concept of image inversion interferometers was a Mach-Zehnder interferometer that contained two lenses in one of its arms, resulting in an image inversion in that arm. While it is generally possible to build such interferometers, they would suffer from the asymmetric design and the use of dispersive components. Nonidentical paths would make it difficult to achieve high contrast interference over a broad range of wavelengths, a paramount requirement for this method to work for fluorescent light. Furthermore, wave-front aberrations introduced by the lenses (either through imperfections or misalignment), would degrade the contrast of interference.

The chromatic aberrations of this system could be avoided by replacing the lens telescope of Fig. 9.1 with one consisting of curved mirrors, as was shown by Weigel et al. [66]. However, in order to also reduce wave-front aberrations we tried to completely avoid the use of non-planar optical elements in our setup.

10.1 Penta-interferometer

The penta-interferometer setup described in this section was devised and built together with Simon Sindbert. The measurements were taken by Simon Sindbert [62].

10.1.1 Setup

Building an image inversion interferometer with only planar surfaces becomes possible by extending the classical Mach-Zehnder interferometer from a two-dimensional plane into three dimensions, by letting the interferometer's paths follow the edges of a cube as shown in Fig. 10.1. This arrangement was inspired by the geometrical phase (Berry phase) and its effect on polarisation rotation [4]. Light coming from the microscope C is split at the first beam splitter BS_1. In both arms the light is first reflected at a penta prism (PP_1 and PP_2) and then at a mirror (M_1 and M_2). The two paths are recombined at the second beam splitter (BS_2). Lenses L_C and L_D refocus the light to generate image planes at the (optional) pinholes PH_C and PH_D before it is detected in the detectors D_C (constructive channel) and D_D (destructive channel). The relative phase of the two arms can be altered by translating mirror M_2 along the direction indicated, using a piezo actuator.

Following this path alone does not yet yield the desired effect, the two recombined images would not be inverted with respect to each other. However, the use of penta prisms instead of ordinary mirrors introduces an additional reflection in each path. In one path this additional reflection leads to a mirroring of the propagated image along its x-axis (PP_1), in the other along its y-axis (PP_2). This leads to the final relative image inversion.

10.1.2 Measurements of the detection PSF

Method As a light source we used a frequency doubled Nd:YAG laser emitting at $\lambda = 532$ nm. This was coupled into a single mode fibre which in turn served as a point source in our experiment. It was imaged with a *Zeiss* objective (2.5×/0.08), the light was collimated using a $f = 75$ mm lens and the fibre was scanned across – or very near – the interferometer's optical axis in steps of 100 nm using a piezo actuator (*PSt 1000/10/100 VS18*, *Piezomechanik GmbH*, Munich, Germany). No pinholes were used and rather than using integrating detectors, such as photomultiplier tubes, both output channels were directed onto the same camera (*Cascade II 512*, *Photometrics*,

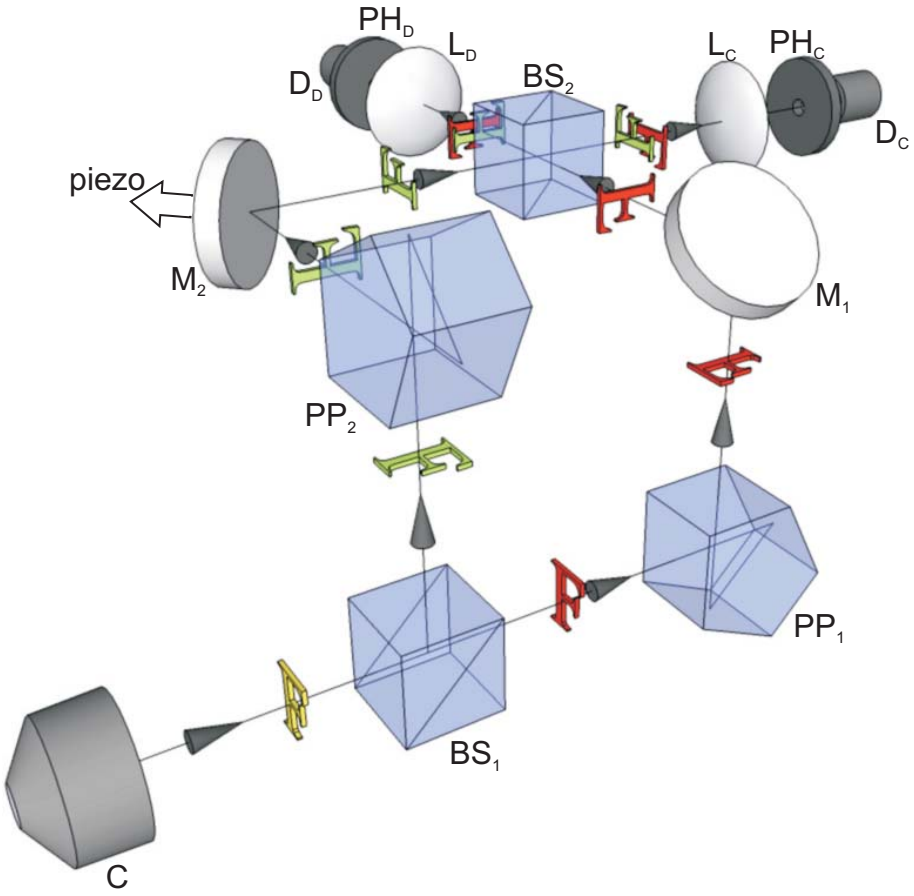
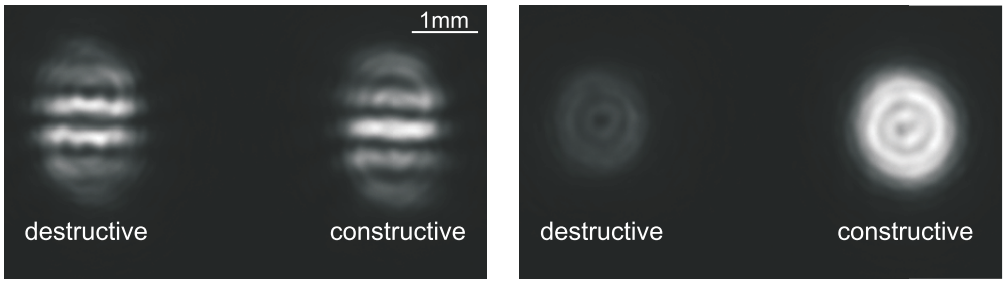


Fig. 10.1: Setup of the penta-interferometer, with a beam path extending in three dimensions. Descanned light coming from the confocal microscope C is split into two paths at the first beam splitter BS_1. After reflections at penta prisms PP_1 and PP_2 and mirrors M_1 and M_2 the two paths are recombined at the second beam splitter BS_2. Lenses L_C and L_D then refocus the light, generating image planes at the (optional) pinholes PH_C and PH_D, before it is detected in the detectors D_C and D_D. The three-dimensional beam path in combination with the penta prisms yields the relative image inversion of the two paths.

Arizona, USA; in non-EM mode), which was placed not in an image plane (except when measuring the wide-field PSF) but in a plane close to the Fourier plane of the system (Fig. 10.2), with the aim of spreading the intensity over a larger area of the CCD chip. This does not affect the results, however, as the integrated intensity in each channel does not depend on which plane a detector is placed in. The wide-field PSF was measured from a CCD image of the fibre (one arm blocked, and focused onto the

CCD) and extrapolated to match the sampling rate to the other curves.

Results The measured intensities are shown in Fig. 10.3b. The modulation of the constructive and destructive signal are clearly visible. The wide-field FWHM resolution was measured to be $d_{\mathrm{wf}} = 4.1$ µm, that of the difference signal of the penta-interferometer was $d_{\mathrm{dif}} = 2.9$ µm. Their ratio of $r_{\mathrm{exp}} = d_{\mathrm{dif}}/d_{\mathrm{wf}} = 0.7$ is close to the theoretical value of $r_{\mathrm{theo}} = 0.68$. Note, that for full-field illumination and under pinhole-free conditions, a regular confocal microscope would have a constant detection PSF and no resolution at all.



(a) (b)

Fig. 10.2: Images recorded for the penta-interferometer. (a) For a point source far away from the optical axis both channels have similar intensities (50%/50% destructive/constructive). The fringes are a result of imaging close to a Fourier plane. (b) On or very near the optical axis the light is predominantly detected in the constructive channel (14%/86% destructive/constructive).

10.1.3 Discussion of penta-interferometer measurements

We were not able to achieve very good contrast for the constructive or destructive interference. The constructive signal rose to around $I_{\mathrm{max}} = 85.75\%$ after offset subtraction of the CCD, while the destructive signal dropped to $I_{\mathrm{min}} = 14.25\%$. The resulting difference signal thus had a contrast of $C_{\mathrm{dif}} = (I_{\mathrm{max}} - I_{\mathrm{min}})/(I_{\mathrm{max}} + I_{\mathrm{min}}) = 71.5\%$. Due to alignment problems we could only show the effect for this type of interferometer using coherent light.

While the penta-interferometer enabled us to demonstrate the resolution improvement for coherent light, it still has its drawbacks: it has a large number of degrees of

freedom, making interferometer alignment very complicated. It is also very difficult to change the path length of the interferometer arms without affecting their propagation direction or lateral position. Furthermore, with the penta prisms there are still dispersive elements in the interferometer, albeit ones that intersect the beam only at right angles; this should guarantee at least a chromatic independence of the path geometry; however, the optical path length will still depend on the wavelength, resulting in a degradation of contrast for broadband light. While these prisms could be replaced with two mirrors per prism, this would further increase the complexity of aligning the interferometer.

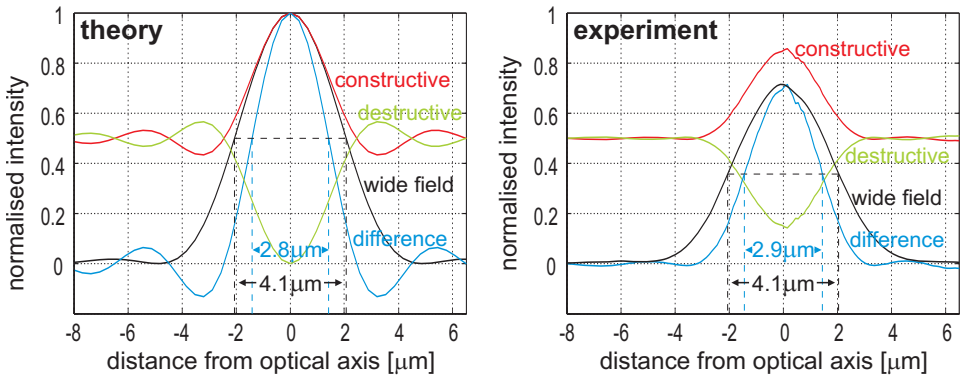


Fig. 10.3: Resulting detection PSFs for a penta-interferometer. (a) shows the theoretical predictions, (b) the experimental results. The constructive output has a maximum signal of 85.8%, the contrast after image subtraction is 71.5%. The FWHM resolution of the difference channel is about 70% that of the wide-field PSF. This is consistent with the theoretically prediced improvement of about 68.4%.

10.2 UZ-interferometer

10.2.1 Setup

To resolve the issues of the penta-interferometer, we devised another interferometer, which is shown in Fig. 10.4. In this much simpler interferometer, descanned light coming from the microscope C is split at the first beam splitter BS_1. The transmitted light is reflected off two mirrors ($M_{Z,1}$, $M_{Z,2}$) to form a Z-shaped path. The reflected

light is also directed via two mirrors ($M_{U,1}$, $M_{U,2}$) and forms a U-shaped path in a plane perpendicular to that of the Z-path. The two paths are recombined at the second beam splitter BS_2. As before, lenses L_C and L_D refocus the light, generating image planes at the (optional) pinholes PH_C and PH_D before detection in the detectors D_C and D_D. As for the penta-interferometer, the relative phase of the two arms can be altered by translating in this case mirror $M_{U,2}$ along the direction indicated, using a piezo actuator. We call this interferometer a UZ-interferometer (UZI) after the shape of the two interferometer paths.

This design resolves all the issues of the earlier one. It is free of dispersive elements (except for the beam splitters) and even needs fewer reflections than the previous setup. Tuning the relative path lengths is made easy by the geometry of the U-path: it forms an optical trombone that can be used to adjust the path length without otherwise altering the path, simply by translating the two mirrors $M_{U,1}$ and $M_{U,2}$ simultaneously in the direction shown.

10.2.2 Measurements of the detection PSF

For the UZI we measured the detection PSF directly, as the system's integrated response to a point-like source in dependence of the source position, as well as indirectly, as a single non-scanned interferometric image of incoherent wide-field illumination.

Direct measurement of the detection PSF

Method

For the direct measurement of the detection PSF we used a continuous wave (CW) laser emitting at $\lambda = 473$ nm (*M-Quadrat*, Germany). This was coupled into a single mode fibre (*460-HP*, *Thorlabs*, Cambridgeshire, UK; nominal $NA_{\text{fibre}} = 0.13$) which in turn served as a point source in our experiment. It was imaged using a *Zeiss* objective (20×/0.5) and tube lens (*Optovar* 1×). The light was then collimated using a lens of $f_{\text{coll}} = 30$ mm focal length and fed into the interferometer. Apart from a large field aperture no pinholes were used and rather than using integrating detectors, such as

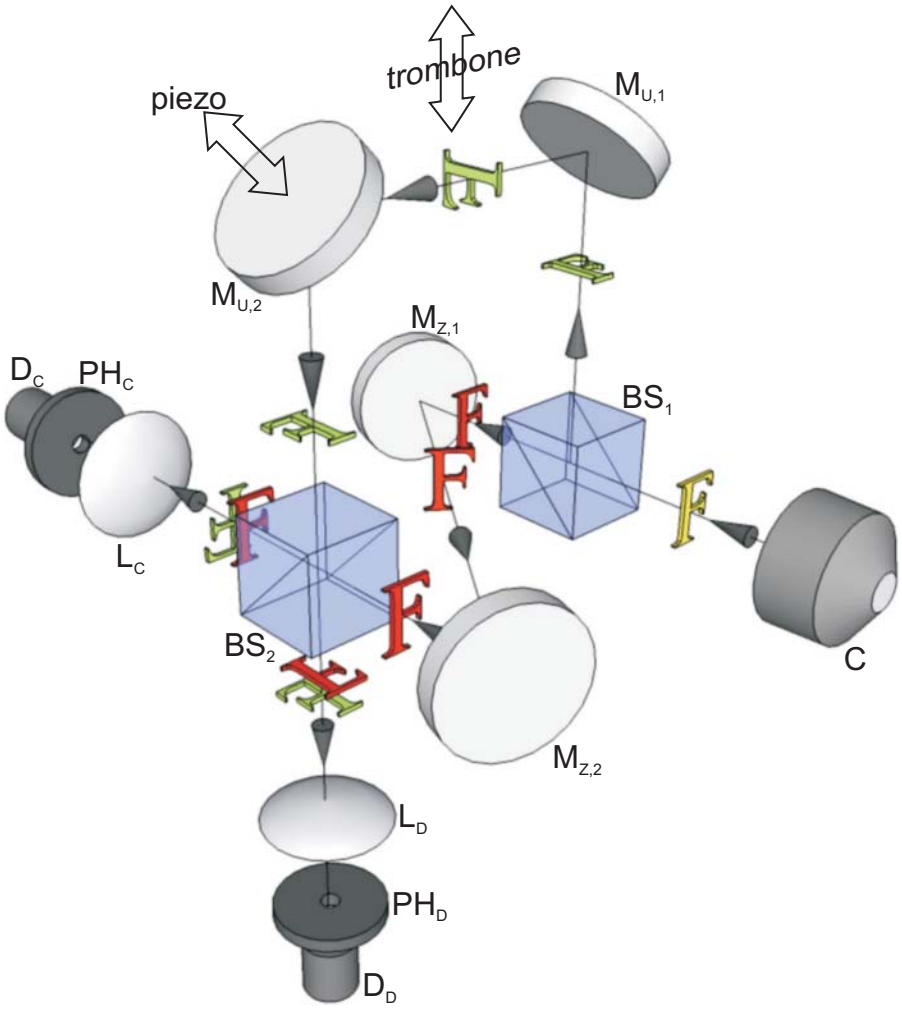


Fig. 10.4: Image inversion in a UZ-interferometer, named after the U- and Z-shaped beam paths. Descanned light coming from the confocal microscope (C) is split at the first beam splitter BS_1. Two mirrors then reflect the light in each arm to form a Z-shaped path ($M_{Z,1}$, $M_{Z,2}$) and a U-shaped path ($M_{U,1}$, $M_{U,2}$), which stand perpendicularly on each other. The two paths are then recombined at the second beam splitter BS_2. Lenses L_C and L_D refocus the light, generating image planes at the (optional) pinholes PH_C and PH_D, before it is detected in the detectors D_C and D_D. The double arrows indicates the translational movement of mirrors $M_{U,1}$ and $M_{U,2}$ to form an optical trombone for adjusting the path length of the interferometer, and the translational movement of mirror $M_{U,2}$ to adjust the relative phase of the interferometer arms.

photomultiplier tubes, both output channels were simultaneously imaged on the same camera (*Imager intense*, *LaVision*, Göttingen, Germany) using a lens of $f_{\mathrm{img}} = 150$ mm focal length and computationally integrated later.

After precise angular alignment the fibre was scanned across the interferometer's optical axis in steps of 193 nm using a piezo translation stage (*Tritor 200/20SG*, *Piezosystems Jena*, Germany).

Results Figure 10.5 shows three frames from this scan (see original publication for a link to the multimedia file of this scan [70]) after offset subtraction. In Fig. 10.5a the fibre is a distance $d = 7.7$ µm away from the inversion axis, where hardly any interference occurs. In Fig. 10.5b the fibre is at $d = 1.5$ µm from the inversion axis, and the two amplitude distributions clearly interfere, leading to more light being detected in the constructive channel (right). Figure 10.5c shows the image for the fibre on the inversion axis. Interference leads to practically all the light being detected in the constructive output, whereas the destructive channel (left) remains dark.

Figure 10.6a shows the integrated signals for the constructive and destructive channel which were normalised with respect to the total integrated intensity detected in both output channels. For the fibre on the inversion axis the integrated intensity in the destructive channel drops to $I_{-,\mathrm{min}} = 0.7\%$. For off-axis points the two outputs have slightly different intensity; this is due to a non-perfect 50/50 splitting of the light in the beam splitters, but can be compensated by doing a scaled subtraction (weights used in this experiment were 0.483 and 0.517) when calculating the difference signal, which is shown in Fig. 10.6b. This figure also shows the wide-field signal, which was measured by blocking one interferometer arm, thus inhibiting interference, and integrating a small area of 5×5 pixels on the camera for each step of the scan. This area was chosen for better SNR and is significantly smaller than the spot size generated by the fibre on the camera (FWHM of 50 pixels) and therefore has a negligible effect on the FWHM of the measured curve.

From Eq. 9.1 and Eq. 9.6 one would expect an interferometric detection PSF with a FWHM that is a factor of $\sqrt{2}$ narrower than the corresponding wide-field PSF [67]. However, this is only true for a homogeneously illuminated back focal plane. For a perfect Gaussian distribution ($a(\vec{r}_{xy}) = \exp\{-\vec{r}_{xy}^{\,2}/(2d_0^2)\}$) Eq. 9.1 yields the same

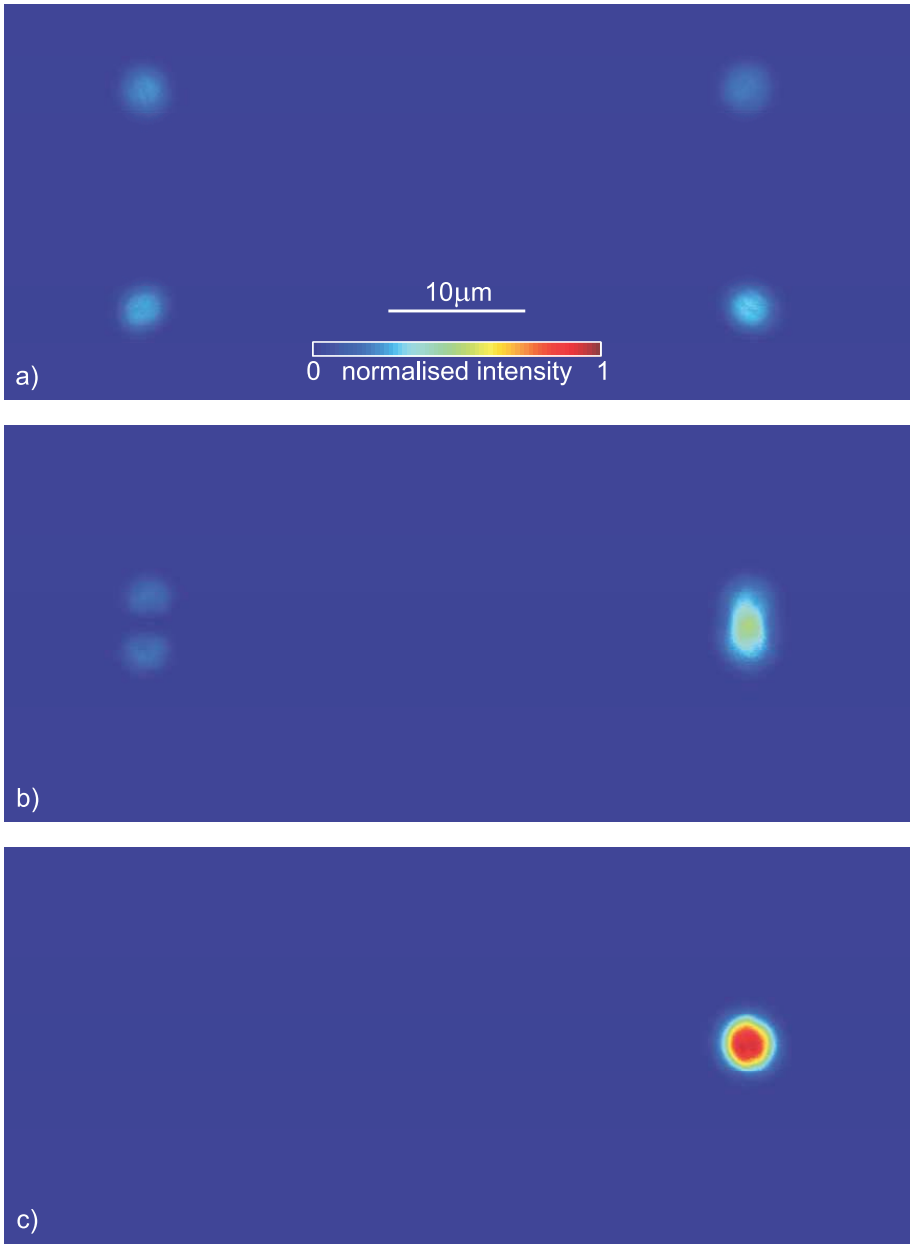


Fig. 10.5: Single frames from fibre scan across the inversion axis (see original publication for a link to the multimedia file of this scan [70]). (a) For a distance of $d = 7.7$ µm from the inversion axis there is hardly any interference and the intensity is divided equally over the two channels. (b) At $d = 1.5$ µm there is interference, leading to more light being detected in the constructive channel (right). (c) When the fibre is located on the inversion axis, interference results in practically all light being detected in the constructive channel, whereas the destructive channel (left) remains dark.

result as the non-interferometric case ($I(\vec{r}_{xy}) = |a(\vec{r}_{xy})|^2$, see appendix B.3). Due to the near-Gaussian mode profile of the light emitted from the fibre with an angular distribution well below the NA of the objective, the measured interferometric detection PSF is only slightly narrower than the corresponding wide-field PSF. This is also the reason for the missing side-lobes in the difference PSF.

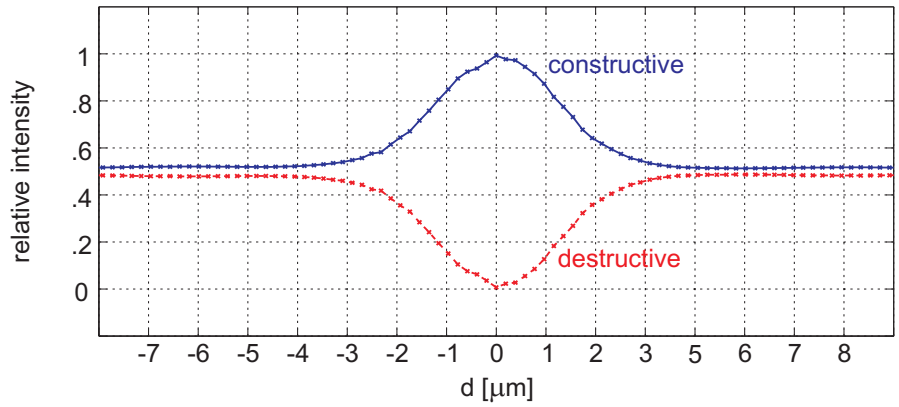
Despite the fact that the difference PSF is only slightly narrower than the wide-field PSF the resolution improvement is still significant, as this detection PSF is achieved under conditions corresponding to full-field illumination and a completely open pinhole (and the corresponding light efficiency). In conventional systems the wide-field PSF could only be achieved for a fully closed pinhole.

Indirect measurement of the detection PSF

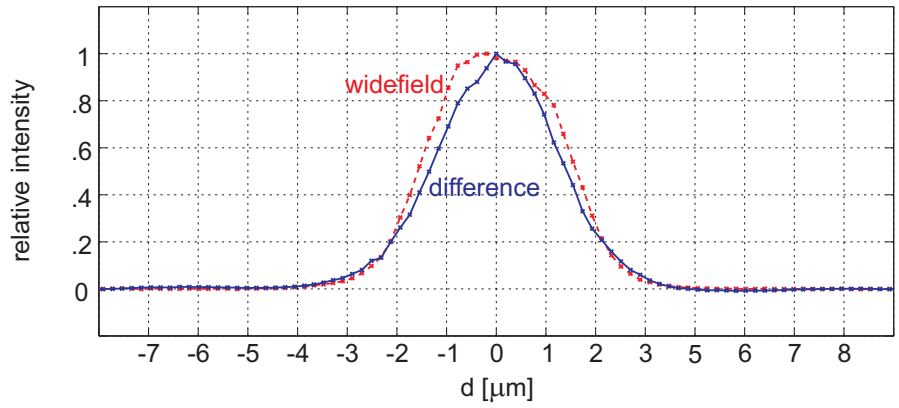
Method For the indirect measurements we took interferometric images of full-field incoherent illumination. The resulting intensity distribution is mathematically equivalent to the detection PSF, which can readily be seen: the intensity distribution generated at the pinhole planes by a single point source is

$$I_{\pm}(\vec{r}_{xy}, \vec{s}_{xy}) = \left|a(\vec{r}_{xy} - \vec{s}_{xy}) \pm a(-\vec{r}_{xy} - \vec{s}_{xy})\right|^2 , \tag{10.1}$$

where $I_{\pm}$ is the intensity in the constructive (+) and destructive (−) channel, a the coherent PSF, $\vec{r}_{xy}$ the pinhole plane coordinate, and $\vec{s}_{xy}$ the source's lateral position in the pinhole plane coordinate system. In a scanning system the detection PSF (full-field illumination, no pinhole, integrating detector) is the detected integrated intensity in dependence of the source position, and we thus have to integrate Eq. 10.1 over the pinhole plane or image coordinate $\vec{r}_{xy}$. On the other hand, to obtain the camera image generated by an incoherently illuminated plane (recorded in the pinhole plane), as we did in this experiment, we have to integrate Eq. 10.1 over all possible source positions $\vec{s}_{xy}$. Interestingly, the two integrations yield the same result; we can therefore indirectly record the constructive and destructive PSF by recording a single image of incoherent



(a)



(b)

Fig. 10.6: Fibre scan across the inversion axis. (a) shows the normalised integrated intensity of the constructive (blue, solid) and the destructive (red, dashed) output. The constructive signal reaches a maximum of 99.3%, the destructive signal drops to 0.7%. The difference in intensity for off-axis points is due to non-perfect 50/50 splitting of the beam splitters. (b) shows the difference signal (the scaled subtraction of constructive and destructive output; blue, solid) and the wide-field signal (red, dashed). The interference signal exhibits hardly any side lobes. This is due to the Gaussian mode profile of the light coming from the fibre.

full-field illumination.

A schematic diagram of the setup is shown in Fig. 10.7. For our measurements we used two light sources: a sodium vapour discharge lamp with two dominant lines at $\lambda_{D,1} = 589.0$ nm and $\lambda_{D,2} = 589.6$ nm; and a white paper screen illuminated by an incandescent white light source (*KL5125*, *A.KRÜSS Optronic*, Hamburg, Germany).

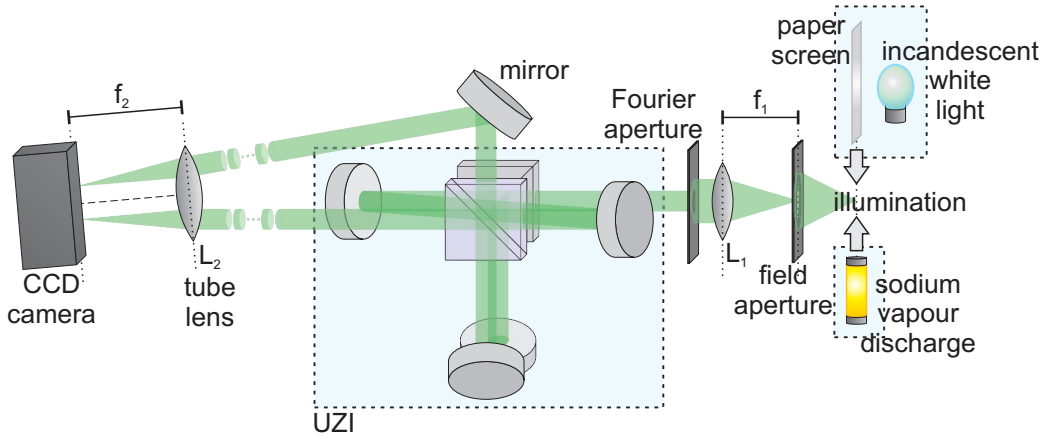


Fig. 10.7: Schematic diagram of the setup used for indirect PSF measurements. Measurements were done for illumination with incandescent white light and sodium vapour discharge.

Both sources homogeneously illuminated a field aperture, which was then imaged using an objective lens (L_1) of $f_1 = 100$ mm focal length before the interferometer and a tube lens (L_2) of $f_2 = 150$ mm to focus both output signals onto the camera. A variable Fourier aperture before the interferometer was used to adjust the numerical aperture of the system, which was kept small at all times (image space $NA = 0.005 \pm 0.001$ derived from a calliper measurement of the aperture), in order to ensure a PSF with large enough FWHM for easy detection. Note that this small NA in image space corresponds to a much larger NA in object space. Both channels were recorded simultaneously using a CCD camera (*Imager intense*, *LaVision*, Göttingen, Germany) and we scanned the path length of one interferometer arm using the actuated mirror, in order to ensure that we measured at the point of equal path length (zero relative phase for all wavelengths). The images were then offset-subtracted and compensated for fluctuations in illumination intensity.

Results

Figure 10.8 shows the constructive signal (a), the destructive signal (b) and the difference in signals (c) recorded for the sodium vapour lamp. We get a maximum construc-

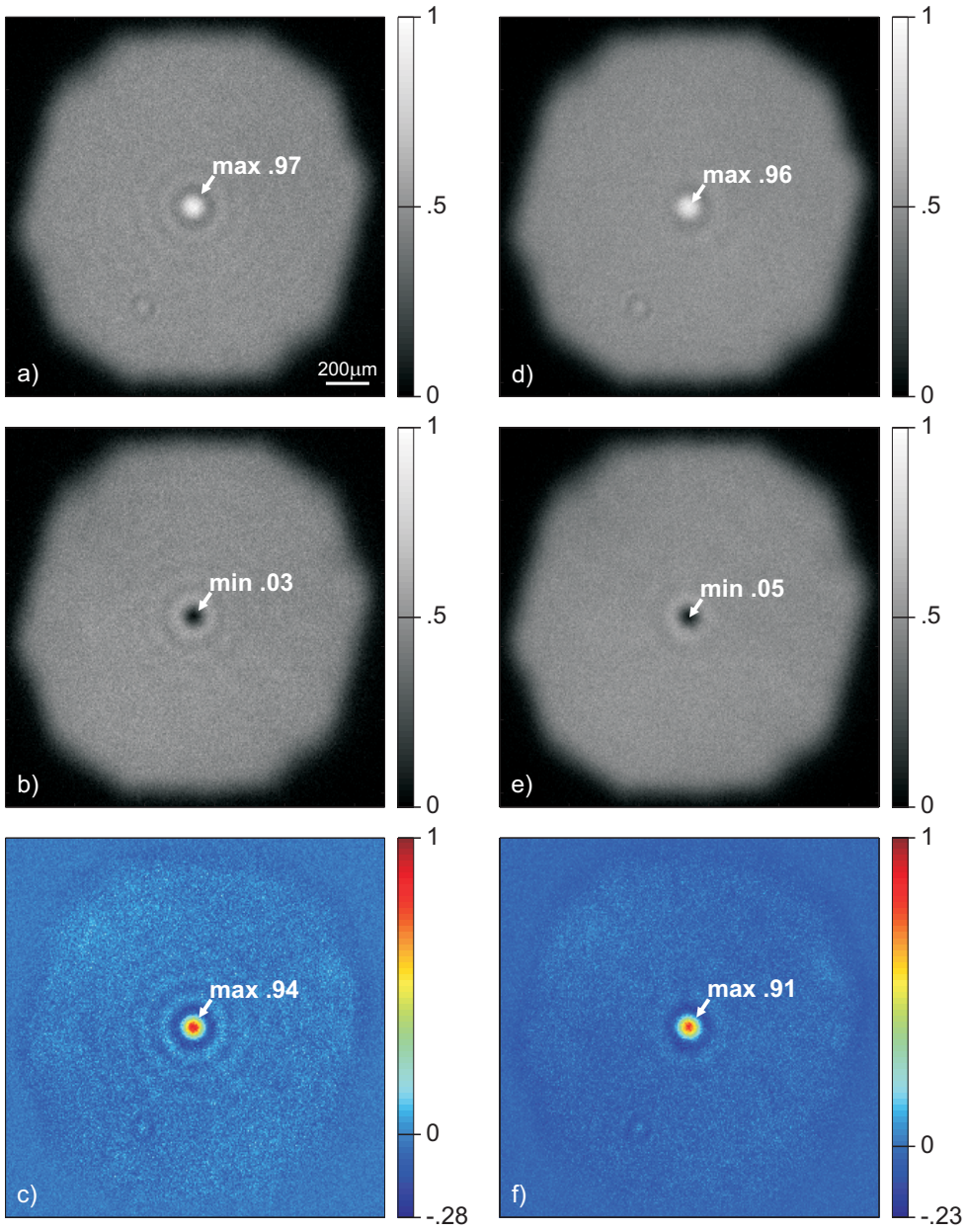


Fig. 10.8: CCD images of the normalised interferometer output. (a)-(c) show the measurements for the sodium vapour lamp, (d)-(f) the white light incandescent lamp. (a) and (d) show the constructive, (b) and (e) the destructive measurement, where the intensity values were normalised so that the non-interfering regions are at $I = 0.5$. The false colour images (c) and (f) show the weighted difference signal.

tive signal of $I_{\max} = 96.8\%$, the destructive signal drops to $I_{\min} = 3.2\%$. The contrast in the difference signal is $C_{\text{dif}} = (I_{\max} - I_{\min})/(I_{\max} + I_{\min}) = 93.6\%$.

Figure 10.8 also shows the constructive signal (d), the destructive signal (e) and the difference in signals (f) recorded for white light. The maximum constructive signal is $I_{\max} = 95.4\%$, the minimum destructive signal is $I_{\min} = 4.6\%$, yielding a contrast of $C_{\text{dif}} = 90.9\%$ in the difference signal.

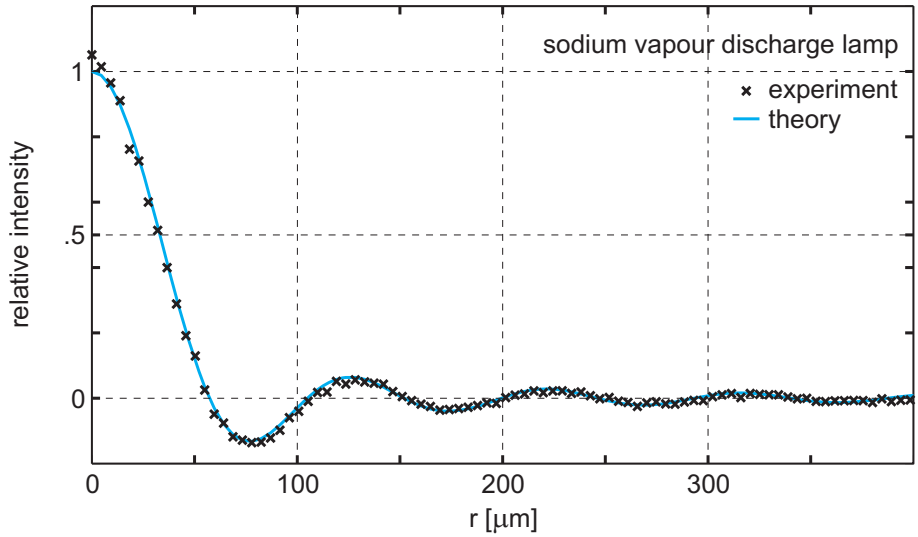
When regarding these interference contrast measurements one has to keep in mind, that the transmission efficiency of the interferometer is only 70% (at 473 nm). We hope that this may be improved through better anti-reflection coatings of the optical elements.

Figure 10.9 shows normalised circular averages of the difference signal in Fig. 10.8c, f. The plot for the sodium vapour discharge lamp (Fig. 10.9a) also shows the theoretical curve, which is $I(r) = 2J_1(2sr)/(2sr)$ [67], where J_1 is the first order Bessel function of the first kind and the parameter s was fitted to match the curve with the experimental data. Note the almost perfect agreement of the measured signal with theory. The FWHM of the curve is 66.2 µm. The relation between FWHM, wavelength λ, numerical aperture NA and magnification m can be expressed as $d_{\text{FWHM}} = 0.356\lambda m/NA$, which gives us $NA_{\text{fit}} = 0.0048$ in good agreement with our coarse estimation of $NA_{\text{estim}} = 0.005 \pm 0.001$.

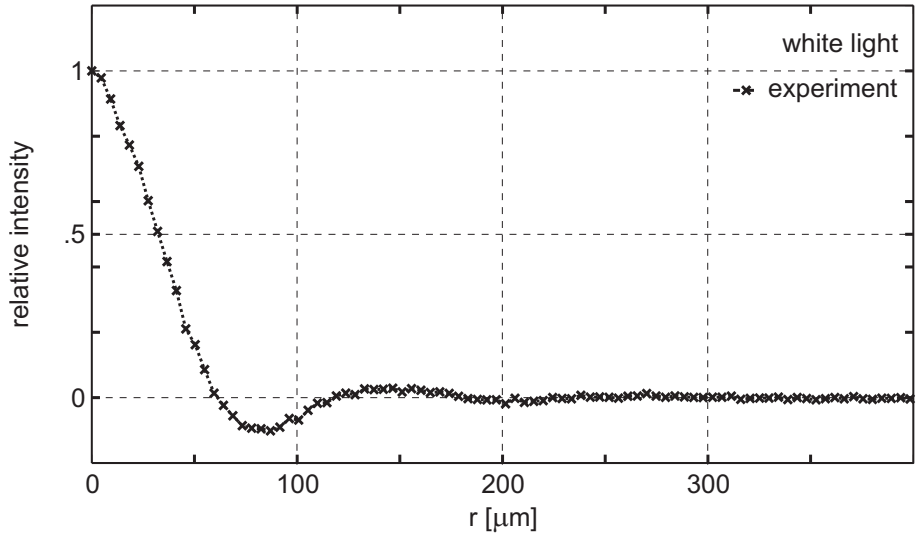
Figure 10.9b shows the equivalent plot for the white light measurements. Here we do not show a theoretical plot, as this would require precise knowledge of the source's spectral composition. Compared to the sodium measurement the white light curve has reduced side lobes. This is due to spectral smearing and may actually be beneficial for applications in microscopy.

10.2.3 Discussion of UZI measurements

We were able to demonstrate the desired interference effect for all three types of light sources. For the penta-interferometer the use of dispersive elements made it virtually impossible to achieve significant interference contrast for anything but laser light. The



(a)



(b)

Fig. 10.9: Circular averages of the difference signals as shown in Fig. 10.8c, f. (a) shows measurements for the sodium vapour discharge lamp, and the theoretically predicted detection PSF (solid line). (b) shows the white light measurements. Due to spectral smearing this curve shows much reduced side lobes compared to the sodium vapour data. A theoretical curve would require precise knowledge of the spectral composition of the light and was therefore omitted.

UZI resolves many of the penta-interferometer's issues: it uses a minimum number of reflections off planar surfaces only and except for the beam splitters uses no dispersive elements. This makes it possible to achieve good interference contrast even for broadband light. Interestingly this setup is also robust against polarisation dependent phase changes possibly occurring in each of the (identical) beam splitter cubes, as the two changes would compensate each other.

In addition to the image inversion the UZI also introduces a geometrical phase difference of π between the two interferometer arms [4], swapping constructive and destructive output at zero optical path difference. This is why even for beam splitters which divide the light in a non-equal ratio this interferometer can theoretically achieve perfect extinction in the destructive channel, as long as both beam splitters show the same splitting ratio.

For broadband light measurements we believe the main current limitation to be dispersion in the beam splitter cubes (*CM1-BS1*, *Thorlabs*, Cambridgeshire, UK): like most commercial beam splitters they have an intentional wedge between entrance and exit surfaces. This is to avoid unwanted interference between Fresnel reflections off those surfaces. However, this wedge character introduces chromatic path variations and prevents us from tuning the interferometer to perfect zero relative phase for all colours simultaneously. A further loss of contrast may arise from unequal dispersion caused by different thickness of the cube beam splitters.

As we were not measuring fluorescence in our indirect PSF measurements, our illumination light may exhibit some partial lateral coherence, potentially making the mathematical description of the detected light more complex. However the deliberate small NA of our imaging system ensured that the observed detection PSF was large compared to the transversal coherence length of the illumination. Possible partial coherence effects of the illumination light could therefore be neglected in the analysis. This was also confirmed by the near perfect agreement of measurement with prediction for the indirect PSF measurement (Fig. 10.9a).

Compared to the indirect PSF measurements the direct measurement (Fig. 10.6)

exhibits practically no side lobes. This is due of the Gaussian mode profile exiting the single mode fibre and the relatively low NA of the fibre (0.13 compared to that of the objective NA of 0.5), which leads to an enhancement of low spatial frequencies.

11. INTERFEROMETRIC IMAGES OF A BIOLOGICAL SAMPLE

11.1 Setup and sample

Setup

For measurements of biological samples we connected the UZI to a *Zeiss Axiovert 200M* microscope. A schematic diagram of the setup is shown in Fig. 11.1.

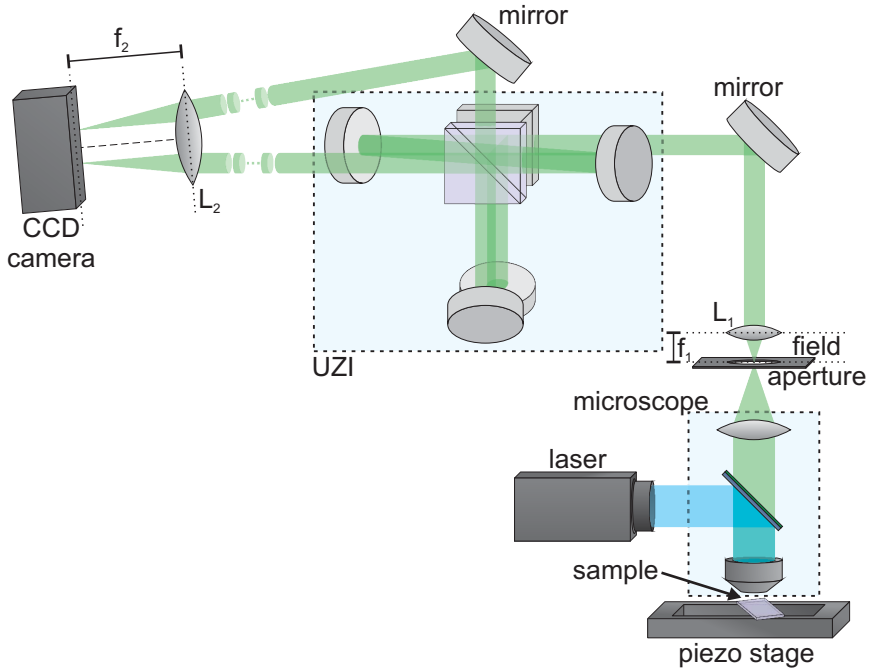


Fig. 11.1: Setup for UZI measurement of biological sample.

For lack of a beam scanning unit, the sample was scanned laterally in steps of approximately 50 nm using a 3-axis Piezo stage (*Tritor 200/20SG*, *Piezosystems Jena*, Germany), which severely limited both scan speed and range. The sample was illu-

minated with a laser at 473 nm, fluorescence was captured by a *Zeiss Plan-Achromat* 63× 1.4 oil immersion objective. Because of the *Axiovert*'s *Optovar* tube lens an additional lens (L_1) of focal length $f_1 = 30$ mm was used to re-collimate the emission light before coupling it into the interferometer. After the interferometer a single lens (L_2) of $f_2 = 200$ mm was used to generate images of both interferometer outputs. Images were then acquired using an electron multiplying CCD (EMCCD) camera (*Cascade II 512*, *Photometrics*, Arizona, USA) with an exposure time of 10 ms per scan position.

Sample

As a sample we used podosomes stained with an *Alexa488*-conjugated antibody embedded in *ProLong* mountant (*Invitrogen*, Paisley, UK), which were kindly provided by James Monypenny of the *Randall Division of Cell and Molecular Biophysics*, *King's College London*.

11.2 Raw images

Pinhole plane images of individual scan position

For each scan position one image of the pinhole plane is acquired. After background subtraction these images are multiplied with a pinhole mask and then integrated to yield the intensity signal at this scan position. This allows us to manipulate the pinhole size after the data acquisition. Figure 11.2 shows an average over 9900 of these scan images prior to integration. The red dashed circle marks a pinhole of 1 AU size. Already in this image a problem with alignment of the interferometer becomes obvious: for perfect co-alignment of the image inversion axis and the illumination axis, the average scan should be symmetric, as the influence of the sample structure is lost in the averaging. However, Fig. 11.2 exhibits an asymmetry, the pattern being elongated and exhibiting stronger side lobes in the vertical direction. Such misalignment may lead to loss in signal-to-noise ratio and even resolution of our technique.

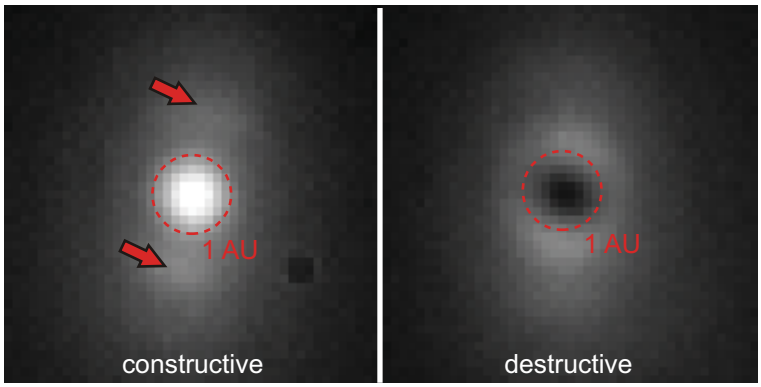


Fig. 11.2: The constructive (left) and destructive (right) outputs are captured on a camera. These scan images are saved individually; confocal-like images are extracted by integrating the intensity inside a pinhole mask (here indicated as a red dashed circle for a pinhole of 1 AU size). For perfectly aligned systems the patterns in this average scan image are expected to be circular symmetric. The vertical elongation and stronger side lobes visible in that direction (indicated by the red arrows) indicate a misalignment of image-inversion axis and illumination axis.

Constructive and destructive output

Integration of the scan images yields a constructive and a destructive image. These can then be subtracted from one another to yield the difference image, or added to yield a confocal image. Figure 11.3 shows these raw images for a pinhole size of 3 AU. Figure 11.3a shows the constructive image, Fig. 11.3b the destructive one. Figure 11.3c shows the sum image, which corresponds to the confocal image (which for a pinhole as large as 3 AU effectively corresponds to a wide-field image), Fig. 11.3d the difference image. While the difference image is relatively noisy, it clearly shows the resolution enhancement over the confocal image for identical pinhole sizes.

The scans exhibit a sudden change in brightness in the upper third of the images. This is due to a lack of synchronisation between the scan start and the start of the illumination. This caused a part of the sample to be illuminated and thus bleached before the actual acquisition of data.

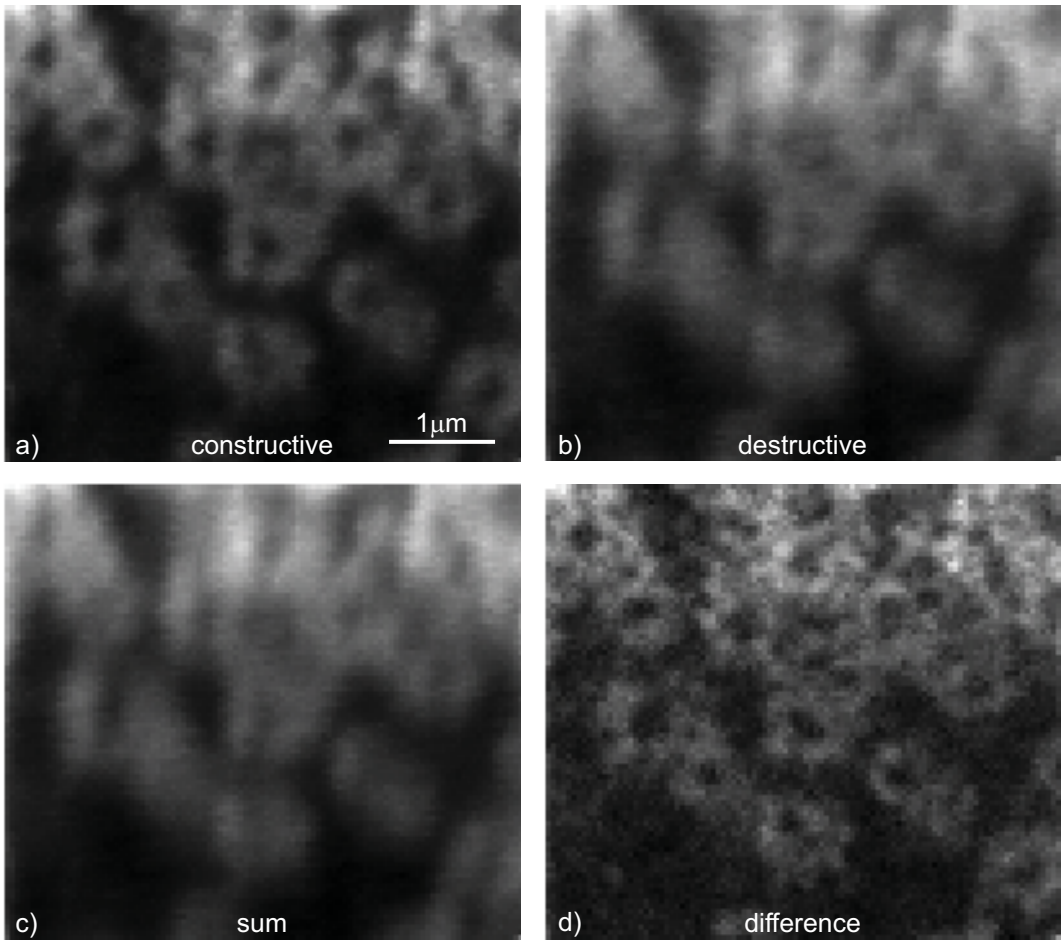


Fig. 11.3: Integration of the individual scan images (Fig. 11.2) yields a constructive (a) and destructive (b) image (here shown for a pinhole of 3 AU size). Summation of these yields the confocal image (c), subtraction the interferometric difference signal (d), for which the resolution enhancement over the confocal case is evident.

11.3 Influence of pinhole size on images

According to Fig. 9.6 we expect the confocal image resolution to degrade with larger pinholes, while the quality of the interferometric difference images improves. And indeed, this behaviour can be observed in Fig. 11.4a-d for the confocal images, and Fig. 11.4e-h for the interferometric ones. As predicted, the images for closed (1/6 AU) pinholes resemble each other (Fig. 11.4a, e). Also in agreement with the predictions in 9.4.3, the SNR of the interferometric images is degraded for very large pinholes (h).

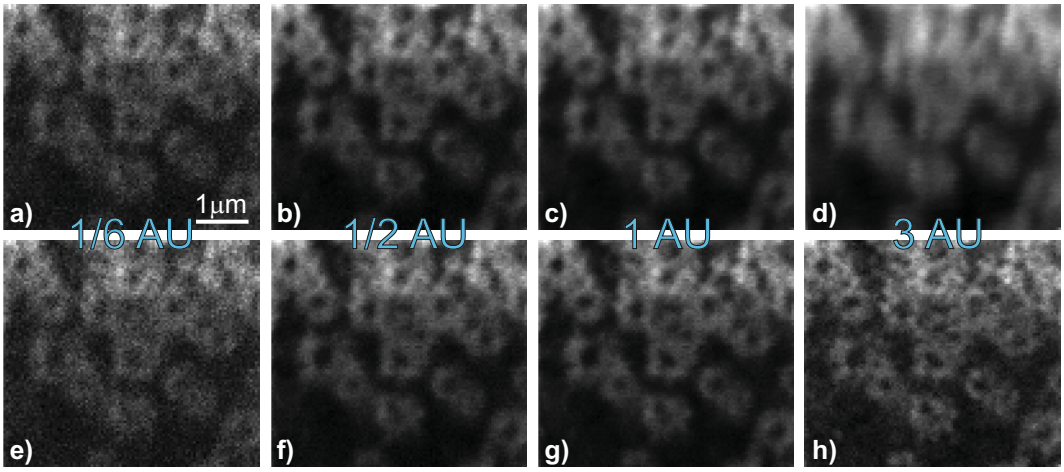


Fig. 11.4: Confocal images (a-d) and interferometric difference images (e-h) for various pinhole sizes. For small pinholes (1/6 AU) the interferometric effect become negligible (e) and the difference image resembles the confocal one (a). For larger pinholes the confocal resolution degrades (b-d) while the interferometric images improve in quality (f, g). For very large pinholes the signal-to-noise levels of the interferometric method actually worsen (h).

For a better comparison of the interferometric technique with convention confocal imaging Fig. 11.5 shows larger versions of an interferometric difference image for a pinhole of 1.5 AU (a) along with confocal images for 1/6 AU (b), 1/2 AU (c) and 1 AU (d). For the interferometric image the pinhole size of 1.5 AU was chosen as a compromise between resolution (better for large pinholes) and SNR (best for pinholes just under 1 AU; Fig. 9.6). Qualitatively, the interferometric image seems better than any of the confocal images. The image quality will be analysed in more detail in the next section.

11.4 Discussion of images

Figure 11.5 suggests a qualitative image improvement for interferometric detection. Figure 11.6 shows an intensity line plot through the line indicated in Fig. 11.5a (light blue) for interferometric and confocal images of various pinhole sizes. Figure 11.6a shows actual intensities. To facilitate the comparison, Fig. 11.6b shows the same curves

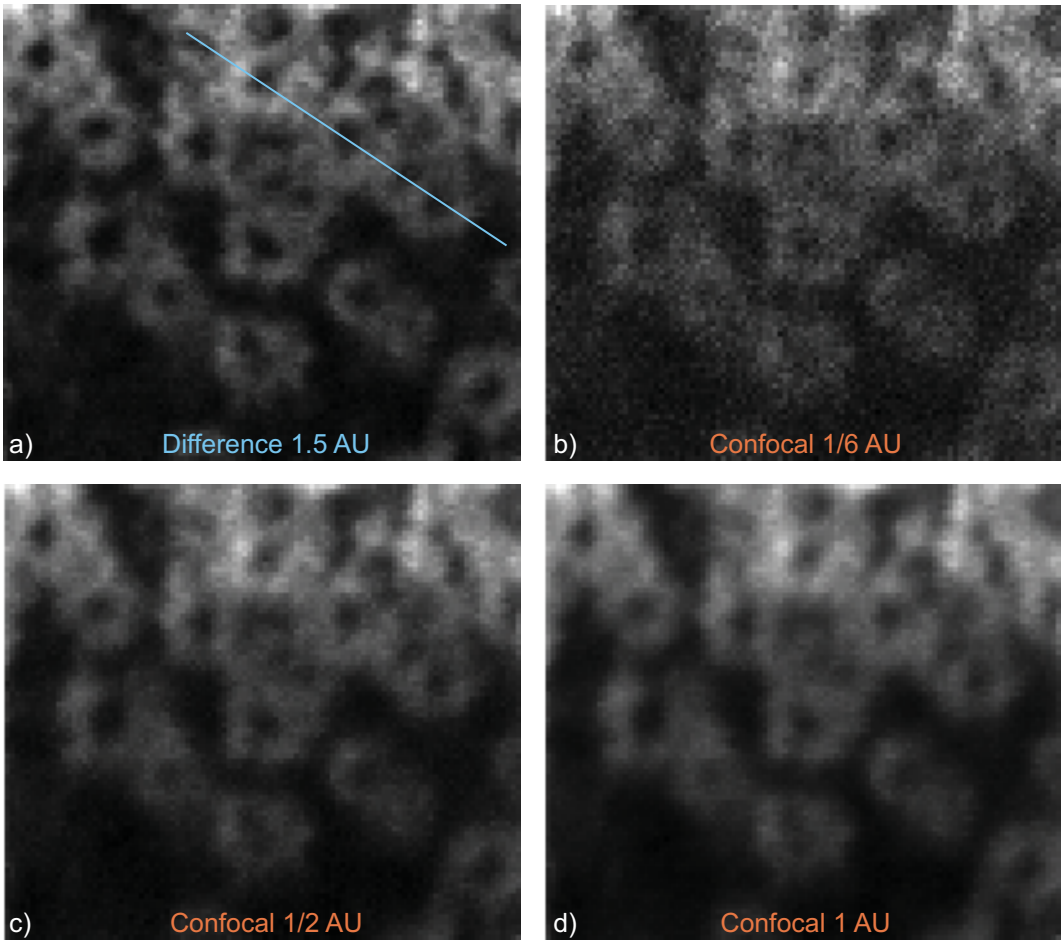


Fig. 11.5: (a) shows an interferometric difference image for a pinhole of 1.5 AU. This pinhole size was chosen as a compromise between good lateral resolution and SNR. The image quality is improved over the confocal images pinhole sizes of 1/6 AU (b), 1/2 AU (c) and 1 AU (d). The light blue line indicates image points used for an intensity line plot (Fig. 11.6). The change in brightness from the top quarter of the image to the bottom part is caused by photo-bleaching.

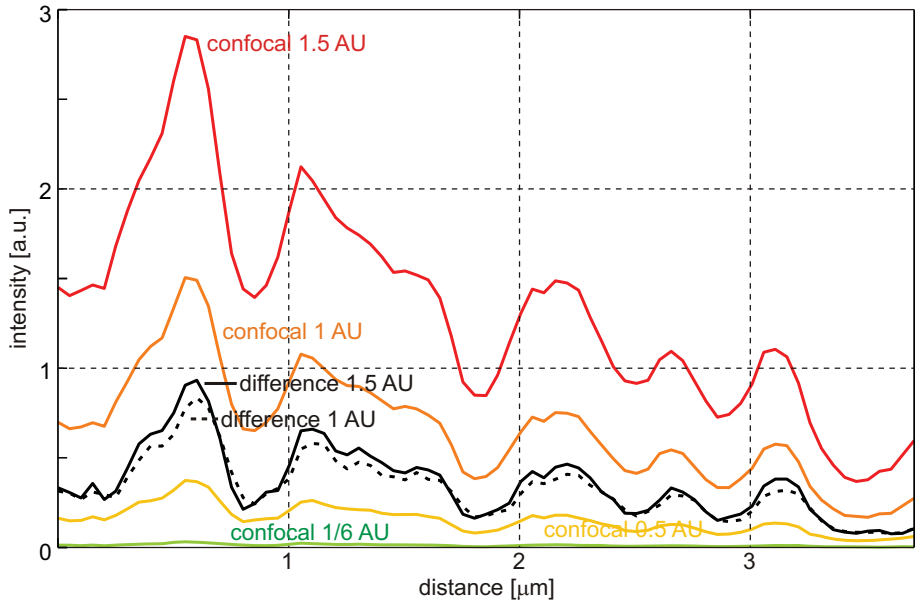
after normalisation to the same mean intensity.

While the interferometric curves (black) have better image contrast and also seems to exhibit fine image detail not present in the confocal images for pinholes larger than 1/2 AU (indicated by grey arrows in Fig. 11.6b), the results are inconclusive regarding resolution. The improved contrast may be due to better optical sectioning and thus reduced background. Similar contrast and image detail is achieved by confocal imaging

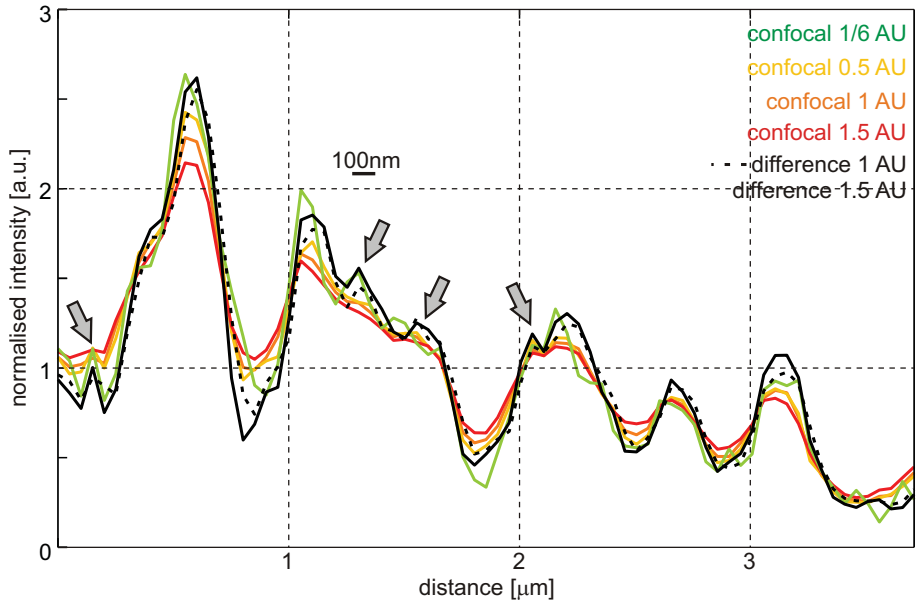
using a small pinhole of 1/6 AU (green). In this case, however, interferometric imaging yield far higher signal (Fig. 11.6a) which results in images of far better SNR (Fig. 11.5).

There are two main reasons for the inconclusiveness of these results: firstly, the podosome sample used in not an ideal sample for resolution testing, and these measurements should eventually be repeated for more appropriate bead samples or other test patterns; and secondly, illumination and interferometer axes are not perfectly co-aligned, as indicated by the asymmetry in Fig 11.2. This leads to a significantly reduced signal after image subtraction and thus reduced SNR. Perfect co-alignment would therefore result in higher interferometric difference signals (Fig. 11.5a). More importantly, however, the misalignment will also result in a degradation of lateral resolution: while for perfect alignment the intensity of illumination and interferometric detection are perfectly correlated regarding source position, a misalignment will degrade this correlation. At positions where the change in intensity of the detection is large, it may be small for the illumination and vice versa.

Better aligned systems may well exhibit improvements better than those already achieved.



(a)



(b)

Fig. 11.6: Intensity plots along the line shown in Fig. 11.6a. (a) shows the true intensities, whereas (b) shows intensities normalised for the same mean intensity. The difference signal (solid black line for 1.5 AU pinhole, dashed black for 1 AU pinhole) shows high contrast and some fine detail (indicated by grey arrows) which is otherwise only visible in the confocal plot for a nearly closed pinhole (i.e. 1/6 AU).

12. IMAGE INVERSION INTERFEROMETRY: OUTLOOK

The theoretical analysis as well as point spread function measurements and first biological images acquired with the UZ-interferometer are very promising indeed. We are optimistic that these results can be improved further by employing beam scanning instead of the much slower stage scanning. This in combination with a more rigid interferometer construction should lead to reduced fluctuations and better images.

Better image quality can also be achieved by combining constructive and destructive channel through weighted averaging in Fourier space rather than a simple subtraction in real space. This has not been possible so far due to the unknown yet significant misalignment of the interferometer and illumination axes, but will be implemented in future versions of the system. If – as in our case – data are recorded by taking individual images for each scan position, image quality can further be improved by looking at the individual pixels rather than simply integrating over a region of the images. This way, each pixel signal will produce a constructive and a destructive scan image with its own PSF, all of which can then again be combined using weighted averaging in Fourier space.

Another possible field of application of image inversion interferometry is imaging with an extended depth of field (EDF) [68]. In this technique samples are illuminated with diffractionless beams, e.g. Bessel beams [10]. While such beams make it possible to acquire a single image for which a very thick section of the sample is in focus, they suffer from very high side-lobes and a slow dropping-off to zero, which leads to hazy images with low contrast. As the detection PSF of the image inversion interferometer is essentially independent of the axial position and therefore also suitable for EDF imaging, it can be combined with Bessel beam illumination to reduce the haze and

improve both contrast and resolution [67]. Using detection in Fourier space the image inversion interferometer could in principle also be used for wide-field EDF imaging [68].

The image inversion interferometer method [27, 58] has been licensed to a company who are planning to commercialise it as an add-on to conventional confocal and two-photon excitation microscopes using the UZ geometry. We are thus hopeful that our technique may have a considerable impact on the field of confocal microscopy.

Appendix

A. STRUCTURED ILLUMINATION MICROSCOPY

A.1 Analysis of sectioning capability of conventional SIM

For a detailed analysis of conventional SIM's [45] sectioning capability let us consider Eq. 3.3,

$$I(\vec{r}_{xy}) = \left| \sum_{n=1}^{3} I_n(\vec{r}_{xy}) \exp\{2\imath n\pi/3\} \right|, \tag{3.3}$$

which using Eq. 3.2,

$$I_n(\vec{r}_{xy}) = \left\{ (L_{\mathrm{illu},n} S) \otimes h_{\mathrm{det}} \right\} (\vec{r})|_{z=0}\,, \; n = \{1,2,3\}, \tag{3.2}$$

can be written as

$$I(\vec{r}_{xy}) = \left| \{ ([\sum_{n=1}^{3} \exp\{2\imath n\pi/3\} L_{\mathrm{illu},n}] S) \otimes h_{\mathrm{det}} \} (\vec{r}) \right|_{z=0}.$$

Using $\sum_{n=1}^{3} \exp\{2\imath n\pi/3\} = 0$ and Eq. 3.1 the inner sum can be simplified to

$$\begin{aligned}
&\sum \exp\{2\imath n\pi/3\} L_{\mathrm{illu},n} \\
=&\sum L/2(\delta(z) \exp\{2\imath n\pi/3\} \cos(\vec{k}_{\mathrm{g}} \cdot \vec{r} + 2n\pi/3)) \otimes h_{\mathrm{illu}}(\vec{r}) \\
=&\sum L/4(\delta(z) \exp\{2\imath n\pi/3\} [\exp\{-2\imath n\pi/3\} \exp\{-\imath \vec{k}_{\mathrm{g}} \cdot \vec{r}\} \\
&+ \exp\{2\imath n\pi/3\} \exp\{\imath \vec{k}_{\mathrm{g}} \cdot \vec{r}\}]) \otimes h_{\mathrm{illu}}(\vec{r}) \\
=&\sum L/4(\delta(z) \exp\{-\imath \vec{k}_{\mathrm{g}} \cdot \vec{r}\}) \otimes h_{\mathrm{illu}}(\vec{r}) \\
=&3/4 \; L \exp\{-\imath \vec{k}_{\mathrm{g}} \cdot \vec{r}\} m(z),
\end{aligned}$$

with the z-dependent modulation function (shown in Fig. 3.1e, f),

$$m(z) = m(\vec{r}) = F_{\vec{k}}^{-1}[\delta(k_x)\delta(k_y)\tilde{h}_{\mathrm{illu}}(k_x - k_{\mathrm{g},x}, k_y - k_{\mathrm{g},y}, k_z)](\vec{r}), \tag{3.6}$$

where F^{-1} denotes the inverse Fourier transform with respect to the direction indicated in the index. The reconstructed image (Eq. 3.3) can then be written as

$$\begin{aligned} I(\vec{r}_{xy}) &= 3/4\ L|[\exp\{-\imath\vec{k}_{\mathrm{g}}\cdot\vec{r}\}m(z)S(\vec{r})] \otimes h_{\mathrm{det}}(\vec{r})|\Big|_{z=0} \\ &= 3/4\ L|\exp\{-\imath\vec{k}_{\mathrm{g}}\cdot\vec{r}\}([m(z)S(\vec{r})] \otimes \{\exp\{\imath\vec{k}_{\mathrm{g}}\cdot\vec{r}\}h_{\mathrm{det}}(\vec{r})\})|\Big|_{z=0} \\ \Rightarrow I(\vec{r}_{xy}) &= 3/4\ L|\exp\{-\imath\vec{k}_{\mathrm{g}}\cdot\vec{r}\}([m(z)S(\vec{r})] \otimes h'_{\mathrm{det}}(\vec{r}))|\Big|_{z=0} \\ &= 3/4\ L|([mS] \otimes h'_{\mathrm{det}})(\vec{r})|\Big|_{z=0}, \end{aligned}$$

with the effective detection PSF $h'_{\mathrm{det}}(\vec{r}) = \exp\{\imath\vec{k}_{\mathrm{g}}\cdot\vec{r}\}h_{\mathrm{det}}(\vec{r}) = F_{\vec{k}}^{-1}[\tilde{h}_{\mathrm{det}}(\vec{k} - \vec{k}_{\mathrm{g}})](\vec{r})$, which contains spatial frequency information filling the missing cone.

A.2 Fluorescence saturation

The fluorescent dyes can be described as two-state systems, with a ground state $|A\rangle$ and excited state $|B\rangle$ [25]. The relative populations of these states are A and B respectively. Transition from $|B\rangle$ to $|A\rangle$ occurs through spontaneous emission at a rate of b, which depends on the fluorescence rate constant. The transition rate from $|A\rangle$ to $|B\rangle$ depends on the illumination intensity L and on a factor a which is inversely proportional to the absorption cross section of the fluorophores.

The change in populations therefore is

$$\begin{aligned} \partial A/\partial t &= -aAL + bB, \\ \partial B/\partial t &= -bB + aAL. \end{aligned}$$

In equilibrium we know that $\partial A/\partial t = -\partial B/\partial t = 0$ and $A + B = 1$, and therefore

$$\begin{aligned} 0 &= -bB + (1-B)aL \\ \Rightarrow bB &= Lab/(b+aL), \end{aligned}$$

which corresponds to the effective illumination.

A.3 Combining components of identical lateral shift

Looking at Eq. 4.3,

$$\tilde{O}_n^{(z_0)}(\vec{k}_{xy}) = L_n \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_n)]\tilde{h}(\vec{k})\, dk_z, \tag{4.3}$$

we see that after the sum projection along k_z any information along this direction is eventually lost. This allows us to rewrite this equation as

$$\tilde{O}_n^{(z_0)}(\vec{k}_{xy}) = L_n \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_{n,xy})]\tilde{h}(\vec{k} + \vec{k}_{n,z})\, dk_z,$$

where $\vec{k}_{n,xy}$ is the projection of $\vec{k}_n$ onto the lateral k_x, k_y-plane, and $\vec{k}_z$ that onto the k_z axis. The sum of any components $\tilde{O}_i^{(z_0)}(\vec{k}_{xy})$ with the same lateral shift vector $\vec{k}_{i,xy}$ can therefore be written as

$$\begin{aligned} \tilde{\Omega}_n^{(z_0)}(\vec{k}_{xy}) &= \sum_i \tilde{O}_i^{(z_0)}(\vec{k}_{xy}) \\ &= \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_{n,xy})] \sum_{i\in\mathbb{I}} \{L_i \tilde{h}(\vec{k} + \vec{k}_{i,z})\}\, dk_z \\ &= \int_{-\infty}^{\infty} [(\exp\{\imath k_z z_0\}\tilde{S}(\vec{k})) \otimes \delta(\vec{k} - \vec{k}_{n,xy})]\tilde{h}_n(\vec{k} + n\vec{k}_{\mathrm{g}})\, dk_z, \\ \tilde{h}_n(\vec{k} + n\vec{k}_{\mathrm{g}})\, dk_z &= \sum_{i\in\mathbb{I}} \{L_i \tilde{h}(\vec{k} + \vec{k}_{i,z})\},\ \mathbb{I} = \{i :\ \vec{k}_{i,xy} = n\vec{k}_{\mathrm{g}}\}. \end{aligned}$$

A.4 Transition from two- to three-dimensional components

Fourier transforming the shifted two-dimensional components $\tilde{\Omega}_n^{(z_0)}(\vec{k}+n\vec{k}_{\mathrm{g}})$ along z_0 yields the three-dimensional components

$$\begin{aligned}
\tilde{\Omega}_n(\vec{k}+\vec{k}_{n,xy}) &= \int_{-\infty}^{\infty} \exp\{-2\pi\imath\ {k'}_z z_0\}\tilde{\Omega}_n^{(z_0)}(\vec{k}_{xy}+\vec{k}_{n,xy})\,dz_0 \\
&= \iint_{-\infty}^{\infty} [\exp\{\imath(k_z-k'-z)z_0\}\tilde{S}(\vec{k})]\tilde{h}_n(\vec{k}+\vec{k}_{n,xy})\,dz_0\,dk_z \\
&= \int_{-\infty}^{\infty} \delta(k_z-k'_z)\tilde{S}(\vec{k})\tilde{h}_n(\vec{k}+\vec{k}_{n,xy})\,dk_z \\
&= \tilde{S}(\vec{k})\tilde{h}_n(\vec{k}+\vec{k}_{n,xy}),
\end{aligned}$$

where in the last line (k_x, k_y, k'_z) was replaced with (k_x, k_y, k_z) to yield the vector $\vec{k}$.

A.5 Weighted averaging

The ideal weights $\tilde{w}_n(\vec{k})$ have to be chosen such that they maximise the signal-to-noise ratio $SNR(\vec{k}) = \tilde{J}_{\mathrm{wa}}(\vec{k})/\sigma_{\mathrm{wa}}(\vec{k})$. This can be done by partially differentiating the SNR with respect to the individual weights and setting the result to zero:

$$\begin{aligned}
&\frac{\partial SNR(\vec{k})}{\partial\tilde{w}_m(\vec{k})} = \frac{\tilde{\Omega}_m(\vec{k}+\vec{k}_{m,xy})}{(\sum_n \tilde{w}_n^2(\vec{k})\sigma_n^2)^{1/2}} - \frac{\tilde{w}_m(\vec{k})\sigma_m^2\sum_n \tilde{w}_n(\vec{k})\tilde{\Omega}_n(\vec{k}+\vec{k}_{n,xy})}{(\sum_n \tilde{w}_n^2(\vec{k})\sigma_n^2)^{3/2}} = 0 \\
\Rightarrow &\tilde{h}_m(\vec{k}+\vec{k}_{m,xy})\sum_n \tilde{w}_n^2(\vec{k})\sigma_n^2 = \tilde{w}_m(\vec{k})\sigma_m^2\sum_n \tilde{w}_n(\vec{k})\tilde{h}_n(\vec{k}+\vec{k}_{n,xy}).
\end{aligned}$$

This yields the weights

$$\tilde{w}_n(\vec{k}) = \tilde{h}_n(\vec{k}+\vec{k}_{n,xy})/\sigma_n^2,\ \forall n. \tag{4.9}$$

A.6 Wiener filter

For the derivation of the Wiener filter function $\tilde{W}(\vec{k})$ we assume the effective imaging PSF and therefore also the effective OTF $\tilde{h}_{\mathrm{wn}}(\vec{k})$ to be real and symmetric. We can thus assume the Wiener function $\tilde{W}(\vec{k})$ to be real and symmetric as well, which will allow us to partially differentiate the mean square error $\varepsilon(\vec{k})$ with respect to $\tilde{W}(\vec{k})$ in

order to find its minimum.

The mean square error is

$$\begin{aligned}\varepsilon(\vec{k}) &= \langle|\tilde{S}(\vec{k}) - \tilde{J}_{\text{wien}}(\vec{k})|^2\rangle \\ &= \langle|\tilde{S}(\vec{k}) - \tilde{J}_{\text{wn}}(\vec{k})\tilde{W}(\vec{k})|^2\rangle \\ &= \langle|\tilde{S}(\vec{k})[1 - \tilde{h}_{\text{wn}}\tilde{W}(\vec{k})] - \tilde{n}(\vec{k})\tilde{W}(\vec{k})|^2\rangle,\end{aligned}$$

where $\tilde{n}(\vec{k})$ is the (white) noise of the Fourier image $\tilde{J}_{\text{wn}}(\vec{k})$. Because all variables are independent of each other, we can write

$$\begin{aligned}\varepsilon(\vec{k}) =&[1 - \tilde{h}_{\text{wn}}(\vec{k})\tilde{W}(\vec{k})]^2\langle|\tilde{S}(\vec{k})|^2\rangle + \tilde{W}^2(\vec{k})\langle|\tilde{n}(\vec{k})|^2\rangle \\ &+ [1 - \tilde{h}_{\text{wn}}(\vec{k})\tilde{W}(\vec{k})]\tilde{W}(\vec{k})\langle\tilde{S}(\vec{k})\tilde{n}^*(\vec{k}) + \tilde{S}^*(\vec{k})\tilde{n}(\vec{k})\rangle \\ =&(1 - \tilde{h}_{\text{wn}}(\vec{k})\tilde{W}(\vec{k}))^2\langle|\tilde{S}(\vec{k})|^2\rangle + \tilde{W}^2(\vec{k})\langle|\tilde{n}(\vec{k})|^2\rangle,\end{aligned}$$

where in the last step we used the fact that for independent noise $\tilde{n}(\vec{k})$ and sample $\tilde{S}(\vec{k})$ follows $\langle\tilde{S}(\vec{k})\tilde{n}^*(\vec{k})\rangle = \langle\tilde{S}^*(\vec{k})\tilde{n}(\vec{k})\rangle = 0$. The assumption that $\tilde{W}(\vec{k})$ is real-valued now allows us to minimise $\varepsilon(\vec{k})$ by differentiating it with respect to $\tilde{W}(\vec{k})$ and setting the result to zero:

$$\begin{aligned}&\frac{\partial\varepsilon(\vec{k})}{\partial\tilde{W}(\vec{k})} = 0 \\ \Rightarrow&(1 - \tilde{h}_{\text{wn}}(\vec{k})\tilde{W}(\vec{k}))\tilde{h}_{\text{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle = \tilde{W}(\vec{k})\langle|\tilde{n}(\vec{k})|^2\rangle \\ \Rightarrow&\tilde{W}(\vec{k}) = \frac{\tilde{h}_{\text{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle}{\tilde{h}^2_{\text{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle + \langle|\tilde{n}(\vec{k})|^2\rangle}.\end{aligned}$$

Inserting this Wiener filter function into the equations for the white-noise recombination of the separated components (Eq. 4.11) yields a Wiener filtered Fourier image

$$\tilde{J}_{\text{wien}}(\vec{k}) = \frac{\sum_n\{\tilde{h}_n(\vec{k} + n\vec{k}_{\text{g}})\tilde{\Omega}_n(\vec{k} + n\vec{k}_{\text{g}})/\sigma_n^2\}}{\sqrt{\sum_n\{\tilde{h}_n^2(\vec{k} + n\vec{k}_{\text{g}})/\sigma_n^2\}}}\frac{\tilde{h}_{\text{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle}{\tilde{h}^2_{\text{wn}}(\vec{k})\langle|\tilde{S}(\vec{k})|^2\rangle + \langle|\tilde{n}(\vec{k})|^2\rangle},$$

which using

$$\tilde{h}_{\mathrm{wn}}(\vec{k}) = \sqrt{\sum_n \{\tilde{h}_n^2(\vec{k} + n\vec{k}_{\mathrm{g}})/\sigma_n^2\}} \tag{4.12}$$

simplifies to

$$\begin{aligned} \tilde{J}_{\mathrm{wien}}(\vec{k}) &= \frac{\langle|\tilde{S}(\vec{k})|^2\rangle \sum_n \{\tilde{h}_n(\vec{k} + n\vec{k}_{\mathrm{g}})\tilde{\Omega}_n(\vec{k} + n\vec{k}_{\mathrm{g}})/\sigma_n^2\}}{\langle|\tilde{S}(\vec{k})|^2\rangle \sum_n \{\tilde{h}_n^2(\vec{k} + n\vec{k}_{\mathrm{g}})/\sigma_n^2\} + \langle|\tilde{n}(\vec{k})|^2\rangle} \\ &= \frac{\langle|\tilde{S}(\vec{k})|^2\rangle \tilde{h}_{wn}^2(\vec{k})\tilde{S}(\vec{k})}{\langle|\tilde{S}(\vec{k})|^2\rangle \tilde{h}_{\mathrm{wn}}^2(\vec{k}) + \langle|\tilde{n}(\vec{k})|^2\rangle}. \end{aligned} \tag{4.14}$$

A.7 Simulation parameters

In chapter 5 several simulations are performed in order to illustrate typical artefacts in light of unknown experimental parameters and our algorithm's ability to cope with such unknowns.

Unless mentioned otherwise, the experimental parameters used in these simulation were:

Microscope Numerical aperture $NA = 1.2$, refractive index of the embedding medium $n = 1.52$, illumination wavelength $\lambda_{\mathrm{illu}} = 488$ nm, detection wavelength $\lambda_{\mathrm{illu}} = 525$ nm, maximum number of photons 3000, camera pixel size (in sample coordinates) $d = 50$ nm, camera background level $b = 500 \pm 30$.

Structured illumination Linear SIM, grating period $g = 500$ nm, three-beam illumination, relative amplitudes of illumination amplitude orders $a_{-1..1} = \{1,\ 0.5,\ 1\}$, number of independent intensity illumination orders $N = 5$, number of directions $D = 3$, directions $\alpha_{1..3} = \{0°, 60°, 120°\}$, number of images per direction $M = 5$, grating positions corresponding to $|\vec{s}_{1..5}| \hat{=} \{0°, 72°, 144°, 216°, 288°\}$.

Sample As as synthetic sample I used a photograph of the ceiling of London's King's Cross railway station taken by Katharina Frassine (Fig. A.1a). The sampling of the sample image corresponds to 12.5 nm per pixel. Figure A.1b shows the corresponding

wide-field image. Both images are 12.8 µm wide.

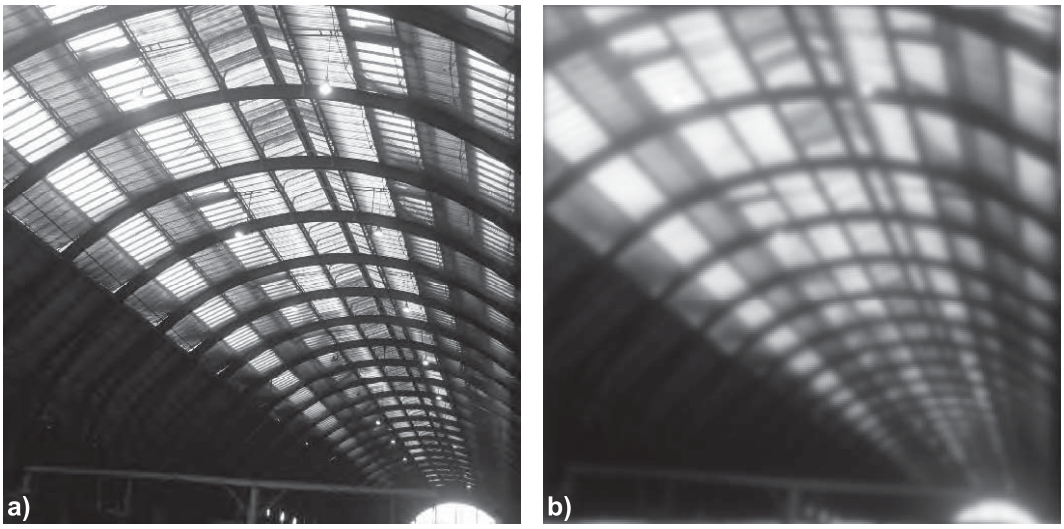


Fig. A.1: As a synthetic sample for the simulation a photograph of the ceiling of London's King's Cross railway station was used (a). The sampling corresponds to 12.5 nm per pixel, the images are 12.8 µm wide. (b) shows the corresponding wide-field image.

A.8 Weighted cross-correlation

Cross-correlations between two (complex) images – and particularly their value at the origin – are used several times in the algorithm, i.e. when optimising the grating's k-vector (5.5), to correct sample drift (5.4, 5.8 and 5.10), in the optimisation of the mixing matrix (5.6) and when determining the global phase (5.9) and the illumination order strengths (5.11).

Unweighted, the cross-correlation image between two (Fourier) images $\tilde{A}(\vec{k})$ and $\tilde{B}(\vec{k})$ is defined as

$$C(\vec{k}) = [\tilde{A} \circledast \tilde{B}](\vec{k}) := \int \tilde{A}(\vec{k}')\tilde{B}^*(\vec{k}+\vec{k}')d^3k.$$

If we are interested in the cross-correlation value of noisy images at one particular $\vec{k} = \vec{k}_0$, a result with better signal-to-noise can be obtained by using a weighted cross-

correlation (WCC) with weights $w(\vec{k})$:

$$C_w = [\tilde{A} \circledast_w \tilde{B}](\vec{k_0}) := \int w(\vec{k'})\tilde{A}(\vec{k'})\tilde{B}^*(\vec{k_0}+\vec{k'})d^3k' / \int w(\vec{k'})d^3k'.$$

In our algorithms we calculate the weights $w(\vec{k})$ by minimising the (absolute square of the) error of the correlation[1]

$$(\Delta C_w)^2 = \frac{\int w^2(\vec{k'})[\langle|\tilde{B}|^2\rangle(\vec{k_0}+\vec{k'})\sigma_A^2(\vec{k'}) + \langle|\tilde{A}|^2\rangle(\vec{k'})\sigma_B^2(\vec{k_0}+\vec{k'})]d^3k'}{(\int w(\vec{k'})d^3k')^2} =: \frac{T_1}{T_2^2},$$

where $\sigma_A^2(\vec{k})$ and $\sigma_B^2(\vec{k})$ are the variances of the two images to be correlated. The error is minimised by differentiating it with respect to the individual weights and setting the result to zero:

$$\begin{aligned}
\frac{(\partial(\Delta C_w)^2)}{\partial w(\vec{k''})} &= \frac{\int 2w(\vec{k'})\delta(\vec{k'}-\vec{k''})[\langle|\tilde{B}|^2\rangle(\vec{k_0}+\vec{k'})\sigma_A^2(\vec{k'}) + \langle|\tilde{A}|^2\rangle(\vec{k'})\sigma_B^2(\vec{k_0}+\vec{k'})]d^3k'}{T_2^2} \\
&\quad - \frac{2T_1\int\delta(\vec{k'}-\vec{k''})d^3k'}{T_2^3} \\
&= \frac{2w(\vec{k''})[\langle|\tilde{B}|^2\rangle(\vec{k_0}+\vec{k''})\sigma_A^2(\vec{k''}) + \langle|\tilde{A}|^2\rangle(\vec{k''})\sigma_B^2(\vec{k_0}+\vec{k''})]}{T_2^2} - \frac{2T_1}{T_2^3} \\
&= 0 \\
\Rightarrow \frac{T_1}{T_2} &= w(\vec{k''})[\langle|\tilde{B}|^2\rangle(\vec{k_0}+\vec{k''})\sigma_A^2(\vec{k''}) + \langle|\tilde{A}|^2\rangle(\vec{k''})\sigma_B^2(\vec{k_0}+\vec{k''})],
\end{aligned}$$

which is fulfilled for weights

$$w_{\vec{k}_0}(\vec{k'}) = \frac{1}{\langle|\tilde{B}|^2\rangle(\vec{k_0}+\vec{k'})\sigma_A^2(\vec{k'}) + \langle|\tilde{A}|^2\rangle(\vec{k'})\sigma_B^2(\vec{k_0}+\vec{k'})}.$$

A.8.1 Problems with weighted cross-correlations

Sample dependent weights Just as for the Wiener filter it is also problematic for weighted cross-correlations to be dependent on sample dependent weights – i.e. weights dependent on the expectation values $\langle|\tilde{A}|^2\rangle$ and $\langle|\tilde{B}|^2\rangle$. These are not known and have

[1] An alternative approach which we are considering for implementation tries to optimise the weights by minimising the *relative* error of correlation, $|\Delta C_w|^2/|C_w|^2$.

to be estimated or replaced. When correlating Fourier SIM components, one approach is to assume a sample of constant expectation value $\langle|\tilde{S}|^2\rangle(\vec{k}) = 1$. As the components correspond to the Fourier transformed sample multiplied with the components' effective OTF, we can then use $\langle|\tilde{A}|^2\rangle = \langle|\tilde{S}|^2\rangle(\vec{k})|\tilde{h}_{n_A}|^2(\vec{k} + n_A\vec{k}_\mathrm{g}) = |\tilde{h}_{n_A}|^2(\vec{k} + n_A\vec{k}_\mathrm{g})$ in the WCCs.

Weights depending on shift $\vec{k}_0$ Several issues limit the use of weighted cross correlations. For one, the weights $w_{\vec{k}_0}$ needed for calculating a WCC $C_w(\vec{k}_0)$ depend on $\vec{k}_0$ and would thus have to be recalculated for each position $\vec{k}_0$ when calculating a whole correlation map, rather than a correlation value for one position. This would make the calculation of a whole WCC image extremely slow.

Mean correlation per weighted volume A more serious problem however is the fact that unlike regular, unweighted CCs the WCC does not calculate the total correlation of two images, but rather the mean correlation density per (weighted) volume element d^3k. This falsely – but significantly – magnifies the correlation values for those $\vec{k}_0$ for which $\int w_{\vec{k}_0}(\vec{k}')d^3k'$ is small, e.g. if two components to be correlated have hardly any overlap.

For these reasons we use WCCs mainly for CC-based parameter optimisations at a fixed $\vec{k}_0$. This way weights do not have to be calculated repeatedly and the $\vec{k}_0$ dependent bias is avoided.

However, in some cases parameter optimisation requires the maximisation of CC-values for varying $\vec{k}_0$, e.g. when trying to find the correct grating vector $\vec{k}_\mathrm{g}$. Here we use theoretical estimations and fitting to narrow down the possible $\vec{k}_0$ to a sub-pixel range prior to correlation. This allows us to then use the same weights for all $\vec{k}_0$ in a good approximation of the true WCC values, and thus avoid both bias and computational cost.

A.8.2 Noise bias in correlations

In a first approximation we assume noise in Fourier images and components to be independent and uncorrelated in $\vec{k}$. When correlating two images, the noise-correlations will therefore cancel and not influence the value of the image correlation.

However, when calculating an auto-correlation of a Fourier image $\tilde{A}(\vec{k})$, the noise correlation will obviously not cancel for $\vec{k}_0 = 0$. This noise contribution, which amounts to $\int w(\vec{k}')\sigma_A^2(\vec{k}')d^3k' / \int w(\vec{k}')d^3k'$ has to be subtracted from the auto-correlation at $\vec{k}_0 = 0$.

A similar effect does not only occur for auto-correlations, but can also play a role when cross-correlating different unshifted, separated Fourier components in SIM. Because all components are linear combinations of the same Fourier images, their noise will not be independent from each other. In 5.6 and appendix A.9 I show how correlations c_{ij} between components $\tilde{\Omega}_i$ and $\tilde{\Omega}_j$ can be calculated from correlations d_{lm} between Fourier images $\tilde{I}_l$ and $\tilde{I}_m$. The noise contribution to correlations between components, c'_{ij}, can be calculated similarly if the noise correlations between the images, d'_{lm}, are known. The noise of Fourier image $\tilde{I}_l$ has a constant (frequency independent) standard deviation σ_l. This and the assumption that noise from different Fourier images is uncorrelated gives us

$$d'_{lm} = \delta_{lm}\sigma_l^2.$$

Using these noise correlations (tensor D') rather than the image correlations (tensor D) in Eq. 5.2 yields the component noise-correlation tensor

$$C' = \bar{M}^{-1}D'\bar{M}^{-1\dagger}$$

and we have to subtract c'_{ij} from the component WCC $[\tilde{\Omega}_i \circledast_w \tilde{\Omega}_j](0)$, i.e. calculate $C - C'$, to get noise-bias free component correlations.

A.9 Speeding up the matrix optimisation

The separated components $\vec{\tilde{\Omega}} = \bar{M}^{-1}\vec{\tilde{I}}$ depend on the mixing matrix $\bar{M}$ and change when the matrix elements are varied. Using correlation values $c_{ij}^{(p)} := \int w(\vec{k}')\tilde{\Omega}_i(\vec{k}')$ $\tilde{\Omega}_j(\vec{k}' - p\vec{k}_\mathrm{g})d^3k'$ in order to iteratively optimise $\bar{M}$ would therefore require a re-calculation of the correlation tensor C after each iteration. For large images this is very time consuming, and for large numbers of iterations will therefore slow down the algorithm considerably.

Fortunately, this can be significantly sped up by rewriting the order correlation tensor as

$$\begin{aligned}
c_{ij}^{(p)} &= [\tilde{\Omega}_i \circledast_w \tilde{\Omega}_j](-p\vec{k}_\mathrm{g}) \\
&= \int w(\vec{k}') \left(\bar{M}^{-1}\vec{\tilde{I}}(\vec{k}')\right)_i \left(\bar{M}^{-1}\vec{\tilde{I}}(\vec{k}' - p\vec{k}_\mathrm{g})\right)_j^* d^3k' \\
&= \int w(\vec{k}') \left(\sum_s \bar{M}_{is}^{-1}\tilde{I}_s(\vec{k}')\right) \left(\sum_t \bar{M}_{jt}^{-1}\tilde{I}_t(\vec{k}' - p\vec{k}_\mathrm{g})\right)^* d^3k' \\
&= \sum_s \sum_t \bar{M}_{is}^{-1}\bar{M}_{jt*}^{-1} \int w(\vec{k}')\tilde{I}_s(\vec{k}')\tilde{I}_t^*(\vec{k}' - p\vec{k}_\mathrm{g})d^3k',
\end{aligned}$$

which after some rearranging can be written as

$$C^{(p)} = \bar{M}^{-1}D^{(p)}\bar{M}^{-1\dagger}, \tag{5.2}$$

where the dagger symbol † denotes the hermitian (or conjugate transpose) of a matrix and

$$\begin{aligned}
d_{st}^{(p)} &= \int w(\vec{k}')\tilde{I}_s(\vec{k}')\tilde{I}_t^*(\vec{k}' - p\vec{k}_\mathrm{g})d^3k' \\
&= [\tilde{I}_s \circledast_w \tilde{I}_t](-p\vec{k}_\mathrm{g})
\end{aligned}$$

are elements of a tensor D containing the weighted cross-correlation values of unshifted Fourier images $\tilde{I}_s(\vec{k})$ and the ones shifted by $p\vec{k}_\mathrm{g}$, $\tilde{I}_t(\vec{k} - p\vec{k}_\mathrm{g})$.

This image correlation tensor D is independent of the mixing matrix $\bar{M}$ and there-

fore does not have to be recalculated after each iteration. In order to calculate the order correlation tensor C the inverse mixing matrix operates directly on D; time consuming cross-correlations have to be calculated only once before the iterative optimisation.

A.10 Complex linear regression

For the derivation of the solution to the complex linear regression [17], we write our two quantities to be compared as a data set $\{(x, y)\}$ with elements

$$x_k = \tilde{\Omega}_i(\vec{k} + i\vec{k}_\mathrm{g})\tilde{h}_j(\vec{k} + j\vec{k}_\mathrm{g}),$$
$$y_k = \tilde{\Omega}_j(\vec{k} + j\vec{k}_\mathrm{g})\tilde{h}_i(\vec{k} + i\vec{k}_\mathrm{g}),$$

with variances $\sigma^2_{x_k}$ and $\sigma^2_{y_k}$ respectively.

Our task is to fit a line $Y = \gamma X$ to this data set which minimises the weighted mean square error of all data points. In relation to a point (X,Y) on that line, a data point (x_k, y_k) has a weighted square error of

$$\varepsilon^2_k(X, Y) = \frac{(x_k - X)^2}{\sigma^2_{x_k}} + \frac{(y_k - Y)^2}{\sigma^2_{y_k}}. \tag{A.1}$$

To find the point on the line that minimises this error, we insert γX for Y, differentiate with respect to X and set the result to zero (doing the same for Y yields the same result):

$$\frac{d\varepsilon^2_k(X, \gamma X)}{dX} = \frac{2(X - x_k)}{\sigma^2_{x_k}} + \frac{2\gamma(\gamma X - y_k)}{\sigma^2_{Y_k}} = 0$$
$$\Rightarrow X = \frac{x_k\sigma^2_{y_k} + \gamma y_k\sigma^2_{x_k}}{\sigma^2_{y_k} + \gamma^2\sigma^2_{x_k}}.$$

Inserting this X and $Y = \gamma X$ into Eq. A.1 yields the minimum square error

$$\varepsilon^2_k = \frac{(x_k(\sigma^2_{y_k} - 1) + \gamma y_k\sigma^2_{x_k})^2}{\sigma^2_{x_k}} + \frac{(x_k\sigma^2_{y_k}/\gamma + y(\sigma^2_{x_k} - 1))^2}{\sigma^2_{y_k}}.$$

The MSE of all data points is proportional to $E^2 = \int \varepsilon_k^2 d^3k$. In order to minimise the MSE, we need to differentiate it with respect to γ and set the result to zero:

$$\begin{aligned} dE^2/d\gamma = d(\int \varepsilon_k^2 d^3k)/d\gamma &= \int \{d\varepsilon_k^2/d\gamma\} d^3k \\ &= \int \{2y_k[x_k(\sigma_{y_k}^2 - 1) + \gamma y_k \sigma_{x_k}^2] - 2x_k/\gamma^2[x_k\sigma_{y_k}^2/\gamma + y_k(\sigma_{x_k}^2 - 1)]\} d^3k \\ &= 0. \end{aligned} \quad \text{(A.2)}$$

This result agrees with [43].

The correct γ is then found by iteratively solving Eq. A.2. Simulations by Mats Gustafsson [17] show this method to give reliable, bias-free results.

Non-iterative solution

As γ is independent of $\vec{k}$, Eq. A.2 can be rewritten as

$$\begin{aligned} 0 &= \gamma^4 A + \gamma^3 B + \gamma C + D; \\ A &:= \int y_k^2 \sigma_{x_k}^2 d^3k, \\ B &:= \int y_k x_k (\sigma_{y_k}^2 - 1) d^3k, \\ C &:= \int x_k y_k (\sigma_{x_k}^2 - 1) d^3k, \\ D &:= \int x_k^2 \sigma_{y_k}^2 d^3k. \end{aligned}$$

After the terms A, B, C and D have been computed, γ could in principle be found analytically.

A.11 Calculating the FWHM resolution from the width of the bead images

Because the imaging PSF is much more elongated along the axial direction than along any lateral direction we can approximate the imaging process as a convolution of a 2D PSF with an axial projection of the sphere onto the focal plane. For a bead of radius

r_{bead} this axial projection has a lateral variance of [33]

$$\sigma^2_{\text{proj}} = \frac{r^2_{\text{bead}} \int_{-1}^{1} \sqrt{1-x^2} x^2 dx}{\int_{-1}^{1} \sqrt{1-x^2} dx} = \frac{1}{4} r^2_{\text{bead}}.$$

A Gaussian with this variance has a FWHM of $d_{\text{bead}} = 2\sqrt{2\ln 2}\sigma_{\text{proj}} = \sqrt{2\ln 2} r_{\text{bead}}$. When convolving two Gaussians their FWHMs add in quadrature and we can thus approximate the PSF's FWHM as

$$d_{\text{PSF}} = \sqrt{d^2_{\text{image}} - d^2_{\text{bead}}},$$

where d_{image} is the FWHM of the Gaussian fitted to the bead image.

With a nominal bead radius of $r_{\text{bead}} = 35.5$ nm and fitted FWHMs of $d_{\text{image,wf}} = 252$ nm for the wide-field image and $d_{\text{image,SIM}} = 105$ nm for SIM this yields an FWHM resolution of $d_{\text{PSF,wf}} = 249$ nm for the wide-field PSF and $d_{\text{PSF,SIM}} = 96$ nm for the SIM PSF.

B. IMAGE INVERSION INTERFEROMETRY

B.1 Derivation of confocal point spread function

Under full-field illumination, a point source at position $\vec{r}$ will generate a light distribution $h_{\mathrm{det}}(\vec{r}' - \vec{r})$ in the image (detector) plane, where $\vec{r}'$ denotes the image coordinate. Under point-scanning illumination, the amount of light emitted by the point source will depend on the brightness of the illumination spot at the location of the point source. This brightness is $h_{\mathrm{illu}}(\vec{r})$, the light distribution in the image plane thus is $h_{\mathrm{illu}}(\vec{r})h_{\mathrm{det}}(\vec{r}' - \vec{r})$. The pinhole function $p(\vec{r}')$ defines, which parts of this distribution will be actually detected by the integrating detector.

The total integrated intensity (i.e. the confocal PSF) thus is

$$\begin{aligned} h_{\mathrm{total}}(\vec{r}) &= \int d^3r' h_{\mathrm{illu}}(\vec{r}) h_{\mathrm{det}}(\vec{r}' - \vec{r}) p(\vec{r}') \\ &= h_{\mathrm{illu}}(\vec{r}) \int d^3r' h'_{\mathrm{det}}(\vec{r} - \vec{r}') p(\vec{r}') \\ \Rightarrow h_{\mathrm{total}}(\vec{r}) &= h_{\mathrm{illu}}(\vec{r})[h'_{det} \otimes p](\vec{r}), \\ \tilde{h}_{\mathrm{total}}(\vec{k}) &= [\tilde{h}_{\mathrm{illu}} \otimes \{\tilde{h}'_{\mathrm{det}} \tilde{p}\}](\vec{k}), \end{aligned} \tag{8.1}$$

where $h'_{\mathrm{det}}(\vec{r}') := h_{det}(-\vec{r}')$ is the mirrored detection PSF.

B.2 Derivation of the interferometric detection point spread function

Assuming full-field illumination and no (or a completely open) pinhole a point source positioned at $\vec{r}_{xy}$ in the focal plane will generate an amplitude distribution $a(\vec{r}'_{xy} - \vec{r}_{xy})$, where $a(\vec{r}'_{xy})$ is the system's 2D in-focus APSF. The corresponding inverted amplitude

distribution is $a(-\vec{r}'_{xy} - \vec{r}_{xy})$. Their interference leads to an integrated intensity signal

$$\begin{aligned} g_{\pm}(\vec{r}_{xy}) &= 1/4 \int \left| a(\vec{r}'_{xy} - \vec{r}_{xy}) \pm a(-\vec{r}'_{xy} - \vec{r}_{xy}) \right|^2 d^2 r'_{xy} \\ &= 1/4 \int \left(a(\vec{r}'_{xy} - \vec{r}_{xy}) \pm a(-\vec{r}'_{xy} - \vec{r}_{xy}) \right) \left(a(\vec{r}'_{xy} - \vec{r}_{xy}) \pm a(-\vec{r}'_{xy} - \vec{r}_{xy}) \right)^* d^2 r'_{xy} \\ &= 1/2 \int \left| a(\vec{r}'_{xy}) \right|^2 d^2 r'_{xy} \pm 1/2 \; \Re\{ \int a(\vec{r}'_{xy}) a^*(-2\vec{r}_{xy} - \vec{r}'_{xy}) d^2 r'_{xy} \} \\ &= 1/2 \pm 1/2 \; \Re\{ [a \otimes a^*](-2\vec{r}_{xy}) \} \\ &= 1/2(1 \pm [a \otimes a^*](-2\vec{r}_{xy})), \end{aligned}$$

in the constructive (+) and destructive (−) channel. $\Re$ denotes the real part of a complex number, and we could perform the last simplifying step because $a \otimes a^*$ is always real only.

B.3 Detection point spread function for Gaussian amplitude distribution

For a Gaussian amplitude distribution $a(\vec{r}) = \exp\{-|\vec{r}|^2/(2d^2)\}$, $\tilde{a}(\vec{k}) = \exp\{-d^2|\vec{k}|^2/2\}$ the detection-only PSF will be

$$\begin{aligned} g_{det}(\vec{r}) &= [a \otimes a^*](-2\vec{r}) && (9.1) \\ \Rightarrow \tilde{g}_{\mathrm{det}}(\vec{k}) &= \tilde{a}(-\vec{k}/2)\tilde{a}^*(\vec{k}/2) && (9.3) \\ &= \exp\{-d^2|\vec{k}|^2/4\} \\ \Rightarrow g_{det}(\vec{r}) &= \exp\{-|\vec{r}|^2/(d^2)\}. \end{aligned}$$

This is identical to the wide-field PSF $h_{\mathrm{wf}}(\vec{r}) = |a|^2(\vec{r}) = \exp\{-|\vec{r}|^2/(d^2)\}$.

C. ABBREVIATIONS AND SYMBOLS

C.1 Abbreviations

2D	Two-dimensional
3D	Three-dimensional
APD	Avalanche photodiode
APSF	Amplitude point spread function
ATF	Amplitude transfer function
BFP	Back focal plane
CC	Cross-correlation
CCD	Charge-coupled device, CCD camera
CW	Continuous wave
EDF	Extended depth of field
EMCCD	Electron multiplying CCD
FPALM	Fluorescence photoactivation localization microscopy
FT	Fourier transform
FWHM	Full width at half maximum
MSE	Mean squared error
NA	Numerical aperture
OTF	Optical transfer function
PALM	Photo-activated localization microscopy
PMT	Photomultiplier tube
PSF	Point spread function
SIM	Structured illumination microscopy
SLM	Spatial light modulator

SNR	Signal-to-noise ratio
STED	Stimulated emission depletion
STORM	Stochastic optical reconstruction microscopy
UZI	UZ-interferometer
WCC	Weighted cross-correlation

C.2 Symbols

$\lvert \square \rvert$	Absolute value of $\square$
a	Amplitude point spread function, APSF
$\tilde{a}$	Amplitude transfer function, ATF
$\square^*$	Complex conjugate of $\square$
$\otimes$	Convolution operator
$\bigotimes_{n=1}^{N} f_n$	Multiple convolution operator; consecutive convolution of all f_n for $n = \{1..N\}$
$\circledast$	Cross-correlation (CC), as in $[A \circledast B](x)$
$\circledast_w$	Weighted cross-correlation (WCC) using weights w; usually taken at only one point, e.g. $[A \circledast_w B](0)$
$\langle \square \rangle$	Expectation value of $\square$
$\tilde{\square}$	Fourier transform of $\square$
$F_x[f(x)](k_x)$	Fourier transform of $f(x)$ with respect to x
$\square^\dagger$	Hermitian transpose of matrix $\square$
$\imath$	Imaginary unit, $\sqrt{-1}$
$\int \square dx$	Integral of $\square$ with respect to x
h	Point spread function, PSF
$\tilde{h}$	Optical transfer function, OTF
$\Re\{C\}$	Real part of a complex number C
$\cdot$	Scalar product of two vectors, as in $\vec{a} \cdot \vec{b}$
σ	Standard deviation

Bibliography

[1] E. Abbe, *Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung*, Archiv fr Mikroskopische Anatomie, 9 (1873), pp. 413–418.

[2] W. B. Amos, J. G. White, and M. Fordham, *Use of confocal imaging in the study of biological structures*, Applied Optics, 26 (1987), pp. 3239–3243.

[3] P. A. Benedetti, V. Evangelista, D. Guidarini, and S. Vestri, *US patent application*, 1996.

[4] M. V. Berry, *Interpreting the anholonomy of coiled light*, Nature, 326 (1987), pp. 277–278.

[5] E. Betzig, G. H. Patterson, R. Sougrat, W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, *Imaging intracellular fluorescent proteins at nanometer resolution*, Science, 313 (2006), pp. 1642–1645.

[6] G. Binnig, C. F. Quate, and C. Gerber, *Atomic force microscope*, Physical Review Letters, 56 (1986), pp. 930–933.

[7] W. Denk, J. Strickler, and W. Webb, *Two-photon laser scanning fluorescence microscopy*, Science, 248 (1990), pp. 73–76.

[8] G. Donnert, J. Keller, R. Medda, M. A. Andrei, S. O. Rizzoliand, R. Lhrmann, R. Jahn, C. Eggeling, and S. W. Hell, *Macromolecular-scale resolution in biological fluorescence microscopy*, in Proceedings of the National Academy of Sciences of the United States of America, vol. 103, 2006, pp. 11440–11445.

[9] C. Dunsby, *Optically sectioned imaging by oblique plane microscopy*, Optics Express, 16 (2008), pp. 20306–20316.

[10] J. Durnin and J. J. J. Miceli, *Diffraction-free beams*, Physical Review Letters, 58 (1987), pp. 1499–1501.

[11] R. Erni, M. D. Rossell, C. Kisielowski, and U. Dahmen, *Atomic-resolution imaging with a sub-50-pm electron probe*, Physical Review Letters, 102 (2009), pp. 096101–096104.

[12] P. P. Ewald, *Introduction to the dynamical theory of X-ray diffraction*, Acta Crystallographica Section A, 25 (1969), pp. 103–108.

[13] J. T. Frohn, H. F. Knapp, and A. Stemmer, *Three-dimensional resolution enhancement in fluorescence microscopy by harmonic excitation*, Optics Letters, 26 (2001), pp. 828–830.

[14] Y. Gluzman, *SV40-transformed simian cells support the replication of early SV40 mutants*, Cell, 23 (1981), pp. 175–182.

[15] M. G. L. Gustafsson, *Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy*, Journal of Microscopy, 198 (2000), pp. 82–87.

[16] ——, *Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution*, in Proceedings of the National Academy of Sciences of the United States of America, vol. 102, 2005, pp. 13081–13086.

[17] ——. private communication, July 2009.

[18] M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, *I5M: 3D widefield light microscopy with better than 100 nm axial resolution*, Journal of Microscopy, 195 (1999), pp. 10–16.

[19] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, *Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination*, Biophysical Journal, 94 (2008), pp. 4957–4970.

[20] R. Heintzmann, *Saturated patterned excitation microscopy with two-dimensional excitation patterns*, Micron, 34 (2003), pp. 283–291.

[21] ——, *Estimating missing information by maximum likelihood deconvolution*, Micron, 38 (2007), pp. 136–144.

[22] R. Heintzmann and C. Cremer, *Laterally modulated excitation microscopy: Improvement of resolution by using a diffraction grating*, in Proceedings of the Society of Photographic Instrumentation Engineers, vol. 3568, 1999, pp. 185–196.

[23] R. Heintzmann and M. G. L. Gustafsson, *Subdiffraction resolution in continuous samples*, Nature Photonics, 3 (2009), pp. 362–364.

[24] R. Heintzmann, Q. S. Hanley, D. Arndt-Jovin, and T. M. Jovin, *A dual path programmable array microscope (PAM): Simultaneous acquisition of conjugate and non-conjugate images*, Journal of Microscopy, 204 (2001), pp. 119–135.

[25] R. Heintzmann, T. M. Jovin, and C. Cremer, *Saturated patterned excitation microscopy - a concept for optical resolution improvement*, Journal of the Optical Society of America A, 19 (2002), pp. 1599–1609.

[26] R. Heintzmann, V. Sarafis, P. Munroe, J. Nailon, Q. S. Hanley, and T. M. Jovin, *Resolution enhancement by subtraction of confocal signals taken at different pinhole sizes*, Micron, 34 (2003), pp. 293–300.

[27] R. Heintzmann and K. Wicker, *UK patent application*, 2007.

[28] S. W. Hell and E. H. K. Stelzer, *Fundamental improvement of resolution with a 4Pi-confocal fluorescence microscope using two-photon excitation*, Optics Communications, 93 (1992), pp. 277–282.

[29] S. W. Hell and J. Wichmann, *Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy*, Optics Letters, 19 (1994), pp. 780–782.

[30] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy*, Biophysical Journal, 91 (2006), pp. 4258–4272.

[31] L. Hirvonen, *Structured illumination microscopy using photoswitchable fluorescent proteins*, PhD thesis, King's College London, UK, 2008.

[32] L. Hirvonen, O. Mandula, K. Wicker, and R. Heintzmann, *Structured illumination microscopy using photoswitchable fluorescent proteins*, in Proceedings of the Society of Photographic Instrumentation Engineers, vol. 6861, 2008, p. 68610L.

[33] L. M. Hirvonen, K. Wicker, O. Mandula, and R. Heintzmann, *Structured illumination microscopy of a living cell*, Eurpean Biophysics Journal, 38 (2009), pp. 807–812.

[34] M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell, *Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins*, in Proceedings of the National Academy of Sciences of the United States of America, vol. 102, 2005, pp. 17565–17569.

[35] J. Huisken, J. Swoger, F. D. Bene, J. Wittbrodt, and E. H. K. Stelzer, *Optical sectioning deep inside live embryos by selective plane illumination microscopy*, Science, 305 (2004), pp. 1007–1009.

[36] A. Ichihara, T. Tanaami, K. Isozaki, Y. Sugiyama, Y. Kosugi, K. Mikuriya, M. Abe, and I. Uemura, *High-speed confocal fluorescence microscopy using a Nipkow scanner with microlenses for 3-D imaging of single fluorescence molecule in real time*, Bioimages, 4 (1996), pp. 57–62.

[37] R. Juškaitis, T. Wilson, M. A. A. Neil, and M. Kozubek, *Efficient real-time confocal microscopy with white light sources*, Nature, 383 (1996), pp. 804–806.

[38] P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. L. Gustafsson, *Super-resolution video microscopy of live cells by structured illumination*, Nature Methods, 6 (2009), pp. 339–342.

[39] K. König, *Clinical multiphoton tomography*, Journal of Biophotonics, 1 (2008), pp. 13–23.

[40] K. Lidke, B. Rieger, T. Jovin, and R. Heintzmann, *Superresolution by localization of quantum dots using blinking statistics*, Optics Express, 13 (2005), pp. 7052–7062.

[41] O. Mandula, *Patterned excitation microscopy*, master's thesis, Charles University Prague, Czech Republic, 2008.

[42] C. W. McCutchen, *Generalized aperture and the three-dimensional diffraction image*, Journal of the Optical Society of America, 54 (1964), pp. 240–244.

[43] C. Moreno, *A least-squares-based method for determining the ratio between two measured quantities*, Measurement Science and Technology, 7 (1996), pp. 137–141.

[44] J. M. Murray, P. A. Appleton, J. R. Swedlow, and J. C. Waters, *Evaluating performance in three-dimensional fluorescence microscopy*, Journal of Microscopy, 228 (2007), pp. 390–405.

[45] M. A. A. Neil, R. Juskaitis, and T. Wilson, *Method of obtaining optical sectioning by using structured light in a conventional microscope*, Optics Letters, 22 (1997), pp. 1905–1907.

[46] D. Oron, E. Tal, and Y. Silberberg, *Scanningless depth-resolved microscopy*, Optics Express, 13 (2005), pp. 1468–1476.

[47] R. Penrose, *A generalized inverse for matrices*, in Proceedings of the Cambridge Philosophical Society, vol. 51, 1955, pp. 406–413.

[48] M. Petràň, M. Hadravsky, M. D. Egger, and R. Galambos, *Tandem-scanning reflected-light microscope*, Journal of the Optical Society of America, 58 (1968), pp. 661–664.

[49] L. Rayleigh, *Investigations in optics with special reference to the spectroscope*, Philosophical Magazine, 8 (1879), p. 261.

[50] B. Richards and E. Wolf, *Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system*, in Proceedings of the Royal Society of London A, vol. 253, 1959, pp. 358–379.

[51] E. Rittweger, K. Y. Han, S. E. Irvine, C. Eggeling, and S. W. Hell, *STED microscopy reveals crystal colour centres with nanometric resolution*, Nature Photonics, 3 (2009), pp. 144–147.

[52] E. Ruska, *The development of the electron microscope and of electron microscopy*, Reviews of Modern Physics, 59 (1987), pp. 627–638.

[53] H. Ruska, B. v. Borries, and E. Ruska, *Die Bedeutung der Übermikroskopie für die Virusforschung*, Archives of Virology, 1 (1939), pp. 155–169.

[54] M. J. Rust, M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*, Nature Methods, 3 (2006), pp. 793–796.

[55] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley-Interscience, MA, 1991.

[56] N. Sandeau and H. Giovannini, *Arrangement of a 4Pi microscope for reducing the confocal detection volume with two-photon excitation*, Optics Communications, 264 (2006), pp. 123–129.

[57] ——, *Increasing the lateral resolution of 4Pi fluorescence microscopes*, Journal of the Optical Society of America A, 23 (2006), pp. 1089–1095.

[58] N. Sandeau, H. Rigneault, and H. Giovannini, *French patent application*, 2006.

[59] S. Santos, K. K. Chu, D. Lim, N. Bozinovic, T. N. Ford, C. Hourtoule, A. C. Bartoo, S. K. Singh, and J. Mertz, *Optically sectioned fluorescence endomicroscopy with hybrid-illumination imaging through a flexible fiber bundle*, Journal of Biomedical Optics, 14 (2009), p. 030502.

[60] L. H. Schaefer, D. Schuster, and J. Schaffer, *Structured illumination microscopy: Artefact analysis and reduction utilizing a parameter optimization approach*, Journal of Microscopy, 216 (2004), pp. 165–174.

[61] H. Siedentopf and R. Zsigmondy, *Über Sichtbarmachung und Größenbestimmung ultramikroskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser*, Journal de Physique Théorique et Appliquée, 2 (1903), pp. 692–702.

[62] S. Sindbert, *Interferometric resolution and efficiency enhancement for scanning fluorescence microscopes*, master's thesis, University of Heidelberg, Germany, 2008.

[63] C. M. Sparrow, *On spectroscopic resolving power*, Astrophysical Journal, 44 (1916), pp. 76–86.

[64] C. Tauber, P. F. G. Rodriguez, V. Loriette, N. Chenouard, B. Dubertret, and J.-C. Olivo-Marin, *Fast and automatic reconstruction of structured illumination microscopy images with multiscale products*, in International Conference on Pattern Recognition, 2008, pp. 1–4.

[65] M. Walde, *Resolution improvement with structured illumination fluorescence microscopy applied to biological research*, bachelor's thesis, Humboldt University Berlin, Germany, 2009.

[66] D. Weigel, H. Babovsky, A. Kiessling, and R. Kowarschik, *Investigation*

of the impulse response of an image inversion interferometer, Optics Communications, 283 (2009), pp. 368–372.

[67] K. Wicker and R. Heintzmann, *Interferometric resolution improvement for confocal microscopes*, Optics Express, 15 (2007), pp. 12206–12216.

[68] ——, *Fluorescence microscopy with extended depth of field*, in Nanoscopy and Multidimensional Optical Fluorescence Microscopy, A. Diaspro, ed., Taylor and Francis, London, UK, 2010, ch. 4.

[69] ——, *Single-shot optical sectioning using polarization-coded structured illumination*, Journal of Optics, Special Issue on Ultrafast Biophotonics (2010). in press.

[70] K. Wicker, S. Sindbert, and R. Heintzmann, *Characterisation of a resolution enhancing image inversion interferometer*, Optics Express, 17 (2009), pp. 15491–15501.